Proteomics of experimental breast cancer

Warmoes, M.O.

2014

document version
Publisher's PDF, also known as Version of record

2

Proteomics of mouse BRCA1-deficient mammary tumors identifies DNA repair proteins with potential diagnostic and prognostic value in human breast cancer

Marc Warmoes[1], Janneke E. Jaspers[2], Thang V. Pham[1], Sander R. Piersma[1], Gideon Oudgenoeg[1], Maarten P.G. Massink[3], Quinten Waisfisz[3], Sven Rottenberg[2], Epie Boven[1], Jos Jonkers[2], Connie R. Jimenez[1]

[1] Oncoproteomics Laboratory, Department of Medical Oncology and [3] Department of Clinical Genetics, VU Medical Center, De Boelelaan 1117, 1081 HV Amsterdam, The Netherlands; [2] Division of Molecular Biology, Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands

This research was originally published in Molecular and Cellular Proteomics, 2012; 11(7):M111.013334. © the American Society for Biochemistry and Molecular Biology

R1
R2
R3
R4
R5
R6
R7
R8
R9
R10
R11
R12
R13
R14
R15
R16
R17
R18
R19
R20
R21
R22
R23
R24
R25
R26
R27
R28
R29
R30
R31
R32
R33
R34
R35
R36
R37
R38
R39

ABSTRACT

Breast cancer 1, early onset (BRCA1) hereditary breast cancer, a type of cancer with defects in the homology-directed DNA repair pathway, would benefit from the identification of proteins for diagnosis, which might also be of potential use as screening, prognostic or predictive markers. Sporadic breast cancers with defects in the BRCA1 pathway might also be diagnosed. We employed proteomics based on one-dimensional gel electrophoresis in combination with nano-LC-MS/MS and spectral counting to compare the protein profiles of mammary tumor tissues of genetic mouse models either deficient or proficient in BRCA1. We identified a total of 3545 proteins, of which 801 were significantly differentially regulated between the BRCA1-deficient and -proficient breast tumors. Pathway and protein complex analysis identified DNA repair and related functions as the major processes associated with the up-regulated proteins in the BRCA1-deficient tumors. In addition, by selecting highly connected nodes, we identified a BRCA1 deficiency signature of 45 proteins that enriches for homology-directed DNA repair deficiency in human gene expression breast cancer data sets. This signature also exhibits prognostic power across multiple data sets, with optimal performance in a data set enriched in tumors deficient in homology-directed DNA repair. In conclusion, by comparing mouse proteomes from BRCA1-proficient and -deficient mammary tumors, we were able to identify several markers associated with BRCA1 deficiency and a prognostic signature for human breast cancer deficient in homology-directed DNA repair.

INTRODUCTION

Breast cancer associated with BRCA1 mutations accounts for 1-2 % of breast cancer cases in the Western world. BRCA1 hereditary breast cancer falls into the molecular subtype of basal-like breast cancer that has a poor prognosis[1]. Sporadic basal-like breast tumors represent ≈10-15 % of all breast carcinomas and comprise many tumors that share key features of BRCA1-associated tumors[2]. A major function of BRCA1 is its role in homology-directed double-strand break repair, a DNA repair mechanism that uses the sister chromatid as a template and therefore repairs double-strand breaks in an error-free manner. Deficiencies in homology-directed DNA repair cause high levels of genomic instability that increase the risk of tumorigenesis[3;4]. Nevertheless, BRCA1 pathway dysfunction may provide an opportunity for therapeutic intervention: preclinical models suggest an increased sensitivity to ionizing radiation and DNA (repair)-targeting agents[3]. In particular, the use of poly(ADP-ribose) polymerase (PARP) inhibitors holds great promise for clinical application. First results from clinical trials support this therapeutic approach for breast cancer[5]. A major clinical challenge remains the identification of patients that are likely to benefit from DNA (repair)-targeting therapy. Global analyses of molecular alterations in sporadic or hereditary breast cancer have mainly used genome and transcriptome profiling methods. These studies yielded a molecular classification of breast cancer[6]. In addition, genomics and transcriptomics studies yielded a number of gene signatures that were prognostic for survival, time to distant metastasis and response to treatment[1;6-15]. Two prognostic signatures, Oncotype DX®[11] and MammaPrint®[,7;16] have currently been registered for clinical use. Recently, Vollebergh *et al*.[13] have found in a retrospective study that a comparative genomic hybridization BRCA1-like classifier predicts the response to intensive platinum-based chemotherapy in patients with high risk breast cancer. The classifier also identifies patients with BRCA1 loss conferred by causes other than mutations. However, the underlying gene products, which would allow for a better understanding of tumor biology and a more practical diagnostic test, remain unknown. To identify patients with BRCA1-like breast cancer, the analysis of tumor proteins may also be useful in selecting patients that might benefit from tailored therapies. Mass-spectrometry based proteomics technologies have matured to the extent that they can now identify and quantify thousands of proteins. Applying these approaches to cancer tissues provides a complementary insight in breast cancer biology and may identify novel diagnostic and prognostic protein profiles and candidate biomarkers. Protein based biomarkers may be of particular advantage in comparison with transcript-based and genomic markers, because they can be measured in routine assays, *e.g.* by antibody-based methods such as immunohistochemistry and ELISA, of which the latter allows for non-invasive testing. In addition, targeted multiplex mass spectrometry is emerging as a novel quantitative strategy for measuring protein signatures in tumor tissues or blood. Proteomic studies of

breast cancer cells and tissues have already shown the potential for candidate biomarkers discovery[17-22]. In a promising pilot study using SELDI-TOF-MS, serum peptide profiles could distinguish women with BRCA1 mutations who developed breast cancer from those who did not (carrier), normal volunteers, and women with sporadic breast cancer with good sensitivity and specificity[23]. To date, no studies employing in depth nano-LC-MS/MS-based proteomics have focused on BRCA1/2-deficient tumor tissues.

In this study, we employed state of the art mass spectrometry-based proteomics to identify proteins associated with BRCA1-deficient breast tumors. To this end, we made use of inbred mouse models that display a minimal amount of genetic variability. As a model for human breast cancers deficient in BRCA1, we analyzed mouse mammary tumors harboring conditional tissue-specific mutations in BRCA1 and p53[24]. The majority of these tumors is highly similar to their human counterpart with respect to histological and molecular characteristics and shows a high level of genomic instability. For comparison, we analyzed two BRCA1-proficient reference tumor models that are genomically stable[25]. We report a BRCA1 deficiency signature based on 45 proteins with DNA repair(-associated) functions that can enrich for homology-directed DNA repair-deficient tumors and identify breast cancer patients with a poor prognosis in various publicly available breast cancer gene expression data sets.

RESULTS

Protein regulations in BRCA1-deficient versus -proficient mouse mammary tumors

For comparative protein profiling, we employed a label-free workflow based on protein fractionation by gel electrophoresis coupled to nano-LC-MS/MS of in-gel digested proteins and spectral counting. Before embarking on a differential analysis, we assessed the reproducibility of our discovery workflow by analyzing three aliquots of a pooled mammary tumor lysate by gel LC-MS/MS (see Supplementary Figure S1A for gel images). In this analysis, 2220 of 2473 proteins (90%, see Supplementary Figure S1B for Venn diagram) were identified in all three replicates with an average CV of 24 % of the normalized spectral counts, indicating a very good reproducibility of the entire discovery workflow.

To identify proteins associated with BRCA1-deficient mammary tumors, we compared the protein expression profiles in five BRCA1-deficient mammary tumors (*Brca1*$^{-/-}$*;p53*$^{-/-}$, carcinoma histology) with five BRCA1-proficient tumors (two *p53*$^{-/-}$ and three *p53*$^{-/-}$*;Cdh1*$^{-/-}$ tumors, all carcinosarcomas). Whereas the carcinomas have an epithelial phenotype, the carcinosarcomas have a mesenchymal phenotype characterized by spindle cell morphology. The protein band patterns obtained after gel electrophoresis of the 10 tumor lysates and Coomassie staining were similar in terms of overall pattern and intensity (Supplementary Figure S2A). A total of 3545 proteins were identified across all 10 samples (Supplementary

Table S1 for all identifications and spectral count data). The number of proteins identified in the BRCA1-deficient tumor samples was 3409, with 1894 proteins identified in all five mammary tumors (Supplementary Figure S2B), indicating acceptable reproducibility of protein identification and quantification across different biological samples. Similar values were obtained for the five BRCA1-proficient tumors (Supplementary Figure S2C)

To obtain a global overview of the data set, we performed unsupervised hierarchical clustering using the normalized spectral count data from all 3545 identified proteins. The BRCA1-deficient and -proficient tumors clustered in separate groups (Figure 1A). The two different BRCA1-proficient tumor types (*p53*$^{-/-}$ and *p53*$^{-/-}$*;Cdh1*$^{-/-}$) did not form two separate groups, but were intermingled, indicating that BRCA1 status and/or histology type were the predominant factors separating the samples. Overlap analysis showed that 338 proteins were uniquely identified in the BRCA1-deficient samples and 136 were uniquely identified in the BRCA1-proficient tumors. Statistical testing[30] revealed 801 proteins with significantly altered abundance in the BRCA1-deficient and -proficient groups ($p < 0.05$) of which 417 were up-regulated in the BRCA1-deficient tumors, whereas 384 were down-regulated. As expected, supervised hierarchical clustering using the 801 differential proteins (Figure 1B) clearly showed two different groups that clustered according to BRCA1/cell type status. See Supplementary Table S2 for all differential proteins. For integration with transcriptomics, we employed the data set of Liu *et al.*[24] containing gene expression data for the same BRCA1-proficient and -deficient mouse models as used in this study, with the exception that most of the tumors in the discovery set were mammary carcinomas. Of the 801 differential proteins, we were able to retrieve mRNA expression data for 565 proteins, of which 429 (76%) had the same direction of differential expression with 201 of these mRNAs (36%) being significantly differentially expressed (Supplementary Table S2).

In summary, a large proportion (23 %) of the mammary tumor tissue proteome is regulated in BRCA1-deficient tumors as compared with proficient tumors. Because a large fraction of proteins showed co-regulation with a transcriptomics data set that only used BRCA1-deficient carcinomas *versus* BRCA1-proficient carcinomas, we conclude that the differential proteins are related mainly to BRCA1 status and only partially to cell type.

Identification of known markers of human BRCA1-deficient breast cancer

Because BRCA1-deficient breast tumors often belong to the highly proliferative basal-like subtype, we examined the abundance of protein markers known in basal-like breast cancer in our data set. In addition, we looked for known markers of human BRCA1 deficiency. Two basal cytokeratin markers (Krt14 and Krt6b) were significantly up-regulated in the BRCA1-deficient mouse tumors (Table 1). The third cytokeratin (Krt5) was up-regulated ($p = 0.066$) with a fold-change of 3.2. ALDH1, a cancer stem cell marker, was exclusively detected in BRCA1-deficient mouse tumors, in accordance with previous findings[35]. PCNA and Ki67,

two well-known proliferation markers[36], were also significantly up-regulated in the BRCA1-deficient mouse tumors. These confirmatory findings underscore the value of these genetic mouse tumor models and the validity of our proteomics approach to identify proteins associated with BRCA1-related or basal-like breast cancers in patients.

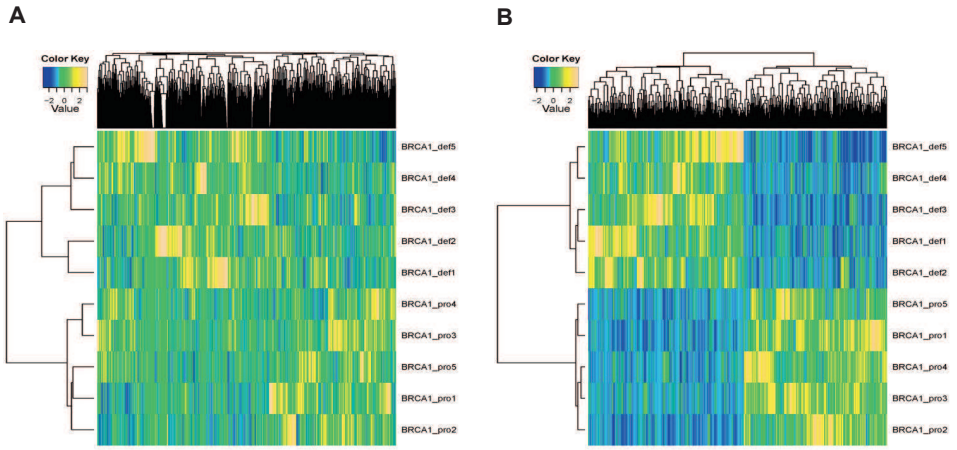


Figure 1. Heat map and cluster analysis using protein expression data from breast tumors of genetic mouse models. A, BRCA1-deficient and -proficient tumors are clustered in separate groups using unsupervised clustering. **B,** supervised clustering clearly separates the BRCA1-deficient tumors from the proficient ones and shows a distinct heat map pattern for up- and down-regulated proteins.

Table 1. Known BRCA1/basal-like and proliferation markers associated with BRCA1 deficiency

Gene name	Basal, proliferation and stem cell markers	Fold change *	*p* value	Marker type
Aldh1a1	Retinal dehydrogenase 1	[a]	0.000471	Stem cell
Krt14	Keratin, type I cytoskeletal 14	6.5	0.023115	Basal
Krt5	Keratin, type II cytoskeletal 5	3.7	0.065672	Basal
Krt6b	Keratin, type II cytoskeletal 6B	[a]	0.016174	Basal
Pcna	Proliferating cell nuclear antigen	1.8	0.001747	Proliferation
Mki67	Ki-67 protein	79.8	1.18E-05	Proliferation

[a] Unique detection in BRCA1-deficient tumors

DNA repair pathways and protein complexes are associated with proteins up-regulated in BRCA1-deficient mammary tumors

To associate biological functions with the differentially expressed BRCA1 deficiency proteins of the mouse mammary tumors, we used the software tool Ingenuity Pathway Analysis. The molecular and cellular functions associated with the up-regulated proteins in BRCA1-

deficient mammary tumors are listed in Table 2, with the number one function identified as DNA Replication, Recombination, and Repair (61 proteins). See Figure 2 for a visualization of the protein network using Ingenuity. The network contains a number of highly connected nodes (*i.e.*, proteins), among which several are well established drug targets (*i.e.*, TOP1, TOP2A, PARP1, and SRC). The top molecular and cellular function associated with the down-regulated proteins was cellular movement (Table 2). The 61 DNA repair proteins up-regulated in BRCA1-deficient mammary tumors were involved in sub-functions like excision repair, chromatin remodeling and modification, double-strand DNA repair and DNA damage response, amongst others. Moreover, canonical pathways associated with the up-regulated proteins in BRCA1-deficient tumors were involved in DNA repair, including ATM signaling, p53 signaling, and role of BRCA1 in DNA damage response (data not shown).

Table 2. Molecular and cellular functions associated with proteins regulated in BRCA1-deficient mammary tumors

Name	p value	No. of proteins
Up-regulated proteins		
DNA Replication, Recombination, and Repair	1.07E-10 – 1.08E-02	61
Cell Cycle	1.19E-09 – 1.08E-02	73
Gene Expression	8.25E-07 – 1.08E-02	73
Cellular Growth and Proliferation	1.90E-06 – 1.08E-02	115
Cellular Development	3.77E-06 – 1.05E-02	39
Down-regulated proteins		
Cellular Movement	4.49E-21 – 1.63E-03	104
Cell Morphology	3.92E-20 – 1.63E-03	85
Cell-To-Cell Signaling and Interaction	6.47E-15 – 1.63E-03	82
Cellular Development	2.26E-13 – 1.70E-03	88
Cellular Function and Maintenance	2.43E-12 – 1.63E-03	69

IPA was used to associate functions to the up- and down-regulated proteins of the BRCA1-deficient mouse tumors. IPA analysis of the up-regulated proteins identified DNA replication, recombination, and repair as the most significant up-regulated molecular and cellular function. Pathway analysis of the down-regulated proteins identified cellular movement as the most significant up-regulated molecular and cellular function.

To identify protein complexes underlying the differential proteins and to further dissect the DNA repair pathways, we employed the COFECO tool[32]. The up-regulated proteins were linked to 53 significant protein complexes (corrected $p < 0.05$, Supplementary Table S3A), of which 44 have a DNA repair(-associated) function (Supplementary Table S3A, highlighted rows). After removing the redundant protein complexes where all members were present in one of the other significant complexes, 29 DNA repair(-associated) complexes were obtained (Supplementary Table S3B).

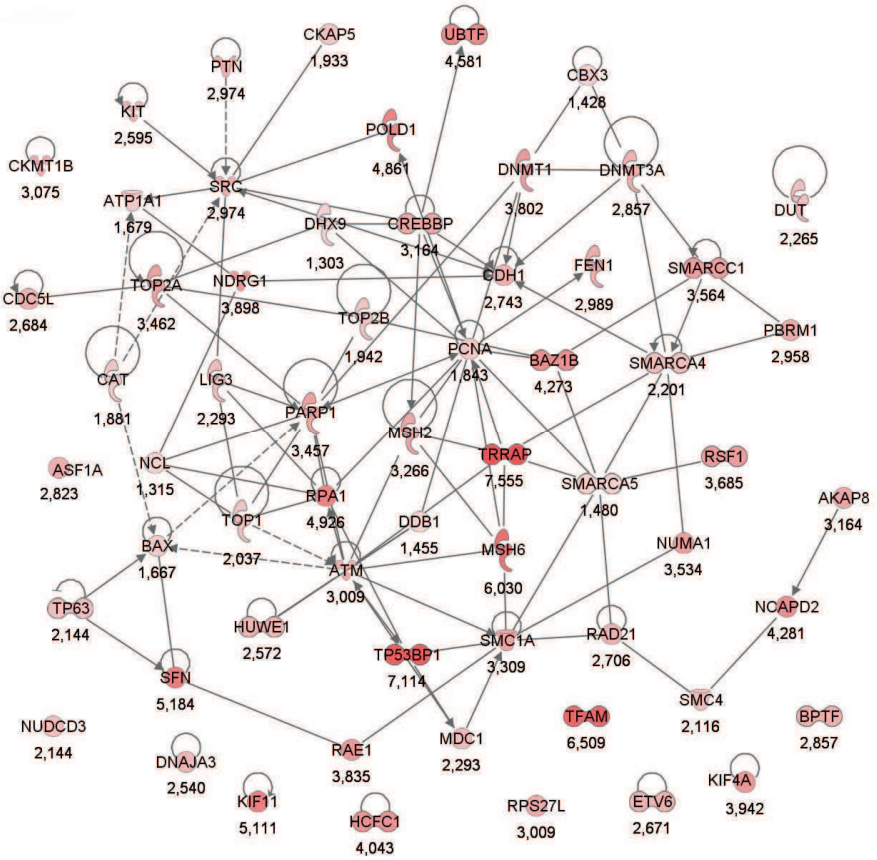


Figure 2. Protein network of significantly up-regulated proteins in BRCA1-deficient tumors that are associated with the top molecular and cellular function DNA replication, recombination, and repair. The nodes represent proteins and the lines between them represent interactions. The dashed lines represent indirect interactions. Color intensity indicates fold change, which is also stated below the nodes.

The DNA repair complexes were involved in chromosome condensation, chromosome cohesion, chromosome remodeling, RNA processing, histone methylation, histone acetylation and the topoisomerase complex, among others. We identified also five complexes which we could not readily link to a physiological process involved in DNA repair. These non-nuclear complexes were involved in integrin cell surface interactions with laminins and collagens. Although these complexes have been implicated in evading apoptosis after DNA damage[37], we did not consider these non-nuclear complexes for further analysis. The down-regulated proteins in BRCA1-deficient tumors were not associated with any DNA repair protein complex in a COFECO analysis, but instead revealed complexes involved in integrin signaling, cytoskeleton regulation and extracellular matrix signaling amongst others (Supplementary Table S3C).

We focused in subsequent analyses on the proteins up-regulated in BRCA1-deficient tumors with a link to DNA repair because we hypothesized that the up-regulation of DNA repair proteins and pathways is linked to BRCA1 status and reflects a compensatory response to the loss of BRCA1 DNA repair function. The 29 nonredundant DNA repair protein complexes associated with the up-regulated proteins in the BRCA1-deficient tumors are visualized in Figure 3 using Cytoscape. It is apparent that many protein complexes have multiple up-regulated members. Examples of DNA repair(-associated) complexes included the BRCA1-associated complex (BASC), involved in double-stranded DNA repair[38], and the condensin I-PARP1-XRCC1 complex with established functions in single-strand DNA repair[39]. In addition, five of seven members of the toposome complex including the drug targets TOP1 and TOP2A were significantly up-regulated[40]. Moreover, many chromatin remodeling complexes, with a wide involvement in different types of DNA repair processes[41], were highly prevalent in our data set. Examples included the WINAC complex, the PBAF complex, the SWI/SNF complex, the GCN5-TRRAP histone acetyl-transferase complex and the DNMT3B histone methylation complex (Supplementary Table S3). Together, the analyses pinpoint a major up-regulation of a broad range of DNA repair/chromatin remodeling pathways and protein complexes in BRCA1-deficient mammary tumors.

Identification of a BRCA1 deficiency DNA repair signature

To identify a protein signature with biological relevance for BRCA1-deficient breast tumors, we reasoned this signature should represent the range of up-regulated DNA repair processes in these tumors and therefore contain selected up-regulated members of each of the 29 nonredundant protein complexes described above. To this end, we selected the most connected up-regulated node in each of the 29 DNA repair protein complexes (Figure 3). This strategy may yield multiple proteins per protein complex, because some nodes show the same level of connectivity. Using this strategy a BRCA1 deficiency signature of 45 proteins was obtained (Table 3). The signature includes PARP1, involved in single-strand base repair; TRRAP, a large adaptor protein involved in histone acetylation; TOP2A, a topoisomerase; SMC1A and SMC4, involved in chromatid cohesion and condensation; BAZ1B and ATM, involved in phosphorylation of H2AX upon DNA damage; and MSH2 and MSH6, involved in mismatch repair.

Up-regulated proteins mapped to human transcripts identify human BRCA1/2-deficient tumors

To investigate the power of the 45 protein BRCA1 deficiency signature in separating BRCA1-deficient and -proficient breast cancers in humans in comparison with all up-regulated proteins, we performed *in silico* analysis using publicly available gene expression data sets (Table 4).

Table 3. List of 45 proteins in BRCA1 deficiency signature

	Protein description	Mouse IPI identifier	Human gene symbol	Fold change	*p* value	Regulation comparison mRNA and protein[a]
R1	Transformation/transcription domain-associated protein	IPI00330902	TRRAP	8,4	<0.001	[c]
R2	Bromodomain adjacent to zinc finger domain protein 1B	IPI00130597	BAZ1B	4,0	<0.001	[d]
R3	Structural maintenance of chromosomes protein 3	IPI00132122	SMC3	2,4	<0.001	[c]
R4	Isoform 2 of Condensin complex subunit 1	IPI00172226	NCAPD2	4,5	<0.001	[d]
R5	Replication protein A 70 kDa DNA-binding subunit	IPI00124520	RPA1	6,3	<0.001	[c]
R6	Isoform 1 of Paired amphipathic helix protein Sin3a	IPI00117932	SIN3A	20,0	<0.001	[d]
R7	Pold1 DNA polymerase	IPI00313515	POLD1	8,5	<0.001	[c]
R8	Activating signal cointegrator 1 complex subunit 3-like 1	IPI00420329	SNRNP200	1,6	0,001	n/a
R9	Structural maintenance of chromosomes protein 1A	IPI00123870	SMC1A	3,2	0,001	[c]
R10	DNA topoisomerase 2-alpha	IPI00122223	TOP2A	4,0	0,001	n/a
R11	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily C member 1	IPI00125662	SMARCC1	3,0	0,001	[e]
R12	DNA topoisomerase 2-beta	IPI00135443	TOP2B	2,1	0,001	[d]
R13	Host cell factor C1	IPI00828490	HCFC1	3,5	0,002	[c]
R14	Proliferating cell nuclear antigen	IPI00113870	PCNA	1,8	0,002	[d]
R15	Rsf1 hepatitis B virus x associated protein	IPI00122845	RSF1	[b]	0,002	[c]
R16	Casein kinase II alpha subunit	IPI00408176	CSNK2A1	2,4	0,002	n/a
R17	Cell division cycle 5-related protein	IPI00284444	CDC5L	3,1	0,002	n/a
R18	DNA topoisomerase 1	IPI00109764	TOP1	2,2	0,003	[c]
R19	Isoform 1 of UDP-N-acetylglucosamine--peptide N-acetylglucosaminyltransferase 110 kDa subunit	IPI00420870	OGT	3,8	0,004	[d]
R20	Isoform 2 of E1A-binding protein p400	IPI00229659	EP400	9,7	0,004	[d]
R21	MutS homolog 6	IPI00310173	MSH6	7,1	0,006	[d]
R22	Isoform 1 of Transcription intermediary factor 1-beta	IPI00312128	TRIM28	1,9	0,008	[c]
R23	Isoform 1 of Splicing factor, arginine/serine-rich 1	IPI00420807	SFRS1	1,4	0,008	n/a
R24	Snf2-related CBP activator protein	IPI00620743	SRCAP	[b]	0,008	n/a
R25	Poly (ADP-ribose) polymerase) 1	IPI00139168	PARP1	3,1	0,008	[c]
R26	SWI/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily a, member 4	IPI00460668	SMARCA4	2,3	0,009	[c]
R27	CREB binding protein	IPI00463549	CREBBP	5,5	0,010	[d]
R28	Serine-protein kinase ATM	IPI00124810	ATM	6,0	0,010	[c]
R29	Double-strand-break repair protein rad21 homolog	IPI00329840	RAD21	3,1	0,014	[d]
R30						
R31						
R32						
R33						
R34						
R35						
R36						
R37						
R38						
R39						

Pre-mRNA-processing-splicing factor 8	IPI00121596	PRPF8	1,5	0,014	[c]
MRG(MORF4-related gene)-binding protein	IPI00720110	C20ORF20	[b]	0,019	[d]
Cleavage stimulation factor 50 kDa subunit	IPI00116747	CSTF1	[b]	0,020	[c]
Metastasis associated 1	IPI00776055	MTA1	[b]	0,020	[c]
DEAD (Asp-Glu-Ala-Asp) box polypeptide 21	IPI00652987	DDX21	2,3	0,020	[c]
Isoform 1 of Heterogeneous nuclear ribonucleoprotein F	IPI00226073	HNRNPF	1,4	0,022	n/a
Nuclear cap-binding protein subunit 1	IPI00458056	NCBP1	1,9	0,028	n/a
SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5	IPI00396739	SMARCA5	1,4	0,030	[c]
DNA mismatch repair protein Msh2	IPI00118158	MSH2	3,0	0,031	[d]
Isoform Long of Splicing factor, arginine/serine-rich 3	IPI00129323	SFRS3	1,4	0,035	[d]
DEAH (Asp-Glu-Ala-His) box polypeptide 9	IPI00339468	DHX9	1,3	0,037	[c]
FACT complex subunit SPT16	IPI00120344	SUPT16H	2,1	0,045	[d]
Cleavage stimulation factor 77 kDa subunit	IPI00116929	CSTF3	6,4	0,046	[d]
Isoform 2 of FACT complex subunit SSRP1	IPI00407571	SSRP1	3,1	0,047	[d]
Structural maintenance of chromosomes protein 4	IPI00229397	SMC4	2,4	0,049	[d]
AT rich interactive domain 1A	IPI00648459	ARID1A	3,8	0,049	[d]

[a] For mRNA regulation
[b] Unique detection in BRCA1-deficient tumors
[c] Significantly up-regulated mRNA
[d] No significant regulation mRNA
[e] Significant opposite regulation mRNA
n/a = no mRNA probe on microarray

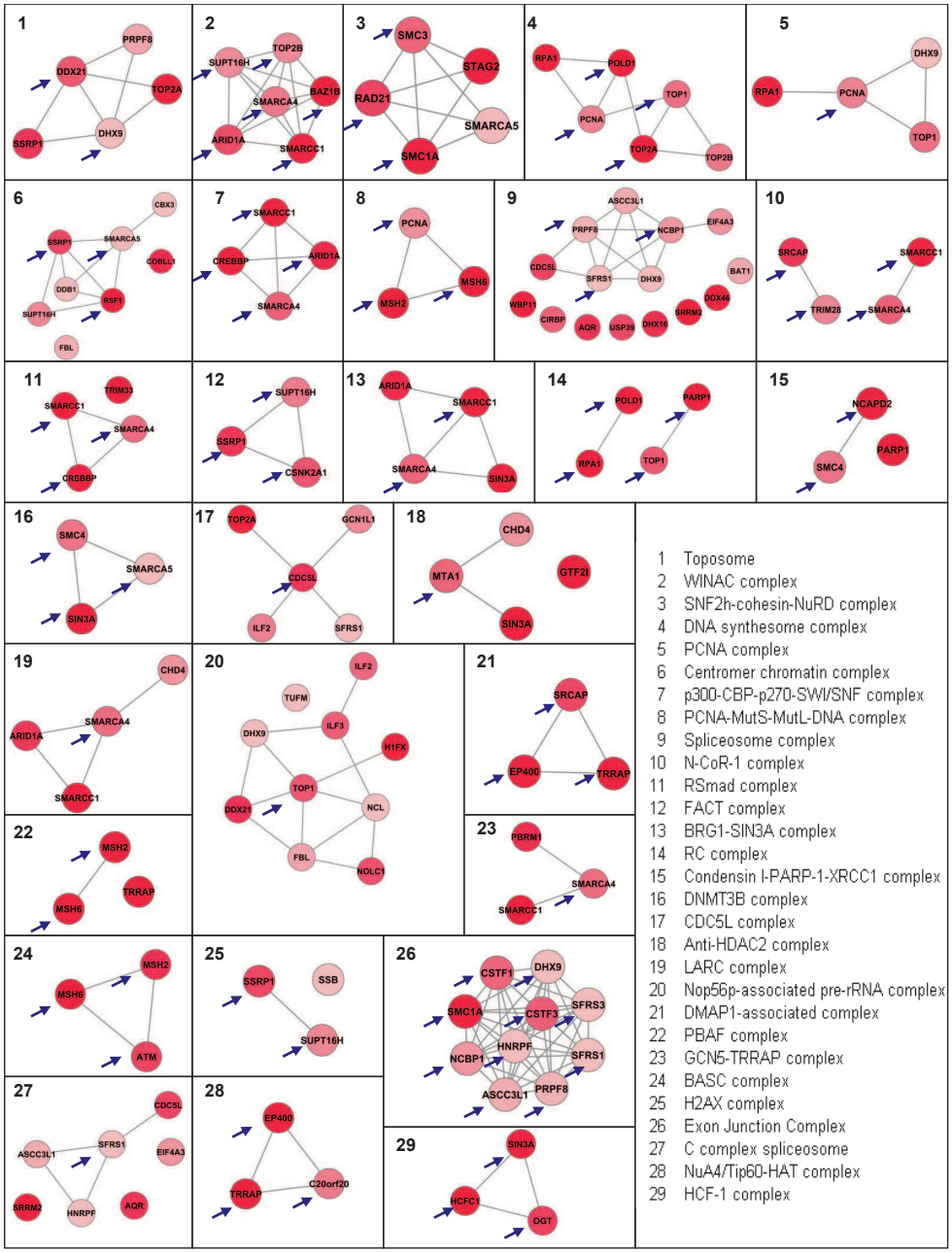


Figure 3. Nonredundant up-regulated DNA repair protein complexes identified by COFECO and visualized using STRING in Cytoscape. The nodes represent proteins and the lines indicate their association as identified in the STRING database. Color intensity is representative of the degree of up-regulation in BRCA1-deficient proteins. Arrows indicate most connected nodes.

We also evaluated the specificity for BRCA2, a gene involved in the same pathway as BRCA1, to examine the ability to identify deficiency in homology-directed DNA repair in general[42]. This is important because of the recent availability of drugs targeting this particular deficiency[43].

We first focused on the Jönsson data set containing 22 *BRCA1*- and 32 *BRCA2*-mutated tumors and other familial and sporadic tumors (Table 4), because this whole genome gene expression data set contained the largest number of *BRCA1/2*-mutated tumors. Hierarchical clustering using all up-regulated proteins showed that the majority of *BRCA1*-mutated tumors were clustered within one branch of the dendrogram, which coincides, as expected, with the basal-like tumors (Figure 4A). The BRCA2 samples were also clustered largely together within the middle branch of the dendrogram. Figure 4B depicts a clustering using the BRCA1 deficiency signature. Here, a large proportion of the BRCA1 and BRCA2 falls within one branch of the dendrogram, making up approximately one-third of the tumors. Thus, the cluster analysis indicates that the 45-protein BRCA1 deficiency signature shows specificity, not only for *BRCA1*-mutated tumors, but also for *BRCA2*-mutated tumors.

Table 4. Description of relevant information of human breast cancer gene expression data sets used for *in silico* validation

	No. of patients	Patient characteristics	Clinical end point	Source
Van de Vijver *et al.*	315	295 sporadic patients	Survival	http://www.rii.com/publications
Van 't Veer *et al.*	20	18 BRCA1 and 2 BRCA2 mutation carriers	n/a	http://www.rii.com/publications
Wang *et al.*	286	286 sporadic patients	Time to distant metastasis	GEO accession GSE2034
Naderi *et al.*	134	134 sporadic patients (120 with survival data)	Survival	Array express accession E-UCON-1
Jönsson *et al.*	359	22 BRCA1, 32 BRCA2, 173 sporadic and 132 non-BRCA1/2 familial patients	Survival	GEO accession GSE22133
Waddell *et al.*	75	19 BRCA1, 30 BRCA2, 1 unknown, and 25 non-BRCA1/2 familial patients	n/a	GEO accession GSE19177

The nearest centroid classification method was employed to characterize more precisely the sensitivity and specificity of the mouse BRCA1 deficiency signature for *BRCA1*- and *BRCA2*-mutated tumors, as well as for the list of all up-regulated proteins. Table 5A reports the classification results on the Jönsson data set with leave-one-out cross-validation. The sensitivities for *BRCA1*-mutated tumors were 77 % and 82 % for the 417 up-regulated proteins and the BRCA1 deficiency signature, respectively. Classification for the combination of BRCA1- and BRCA2-mutated tumors yielded a similar performance: 83 % sensitivity for all up-regulated proteins and 81 % for the BRCA1 deficiency signature.

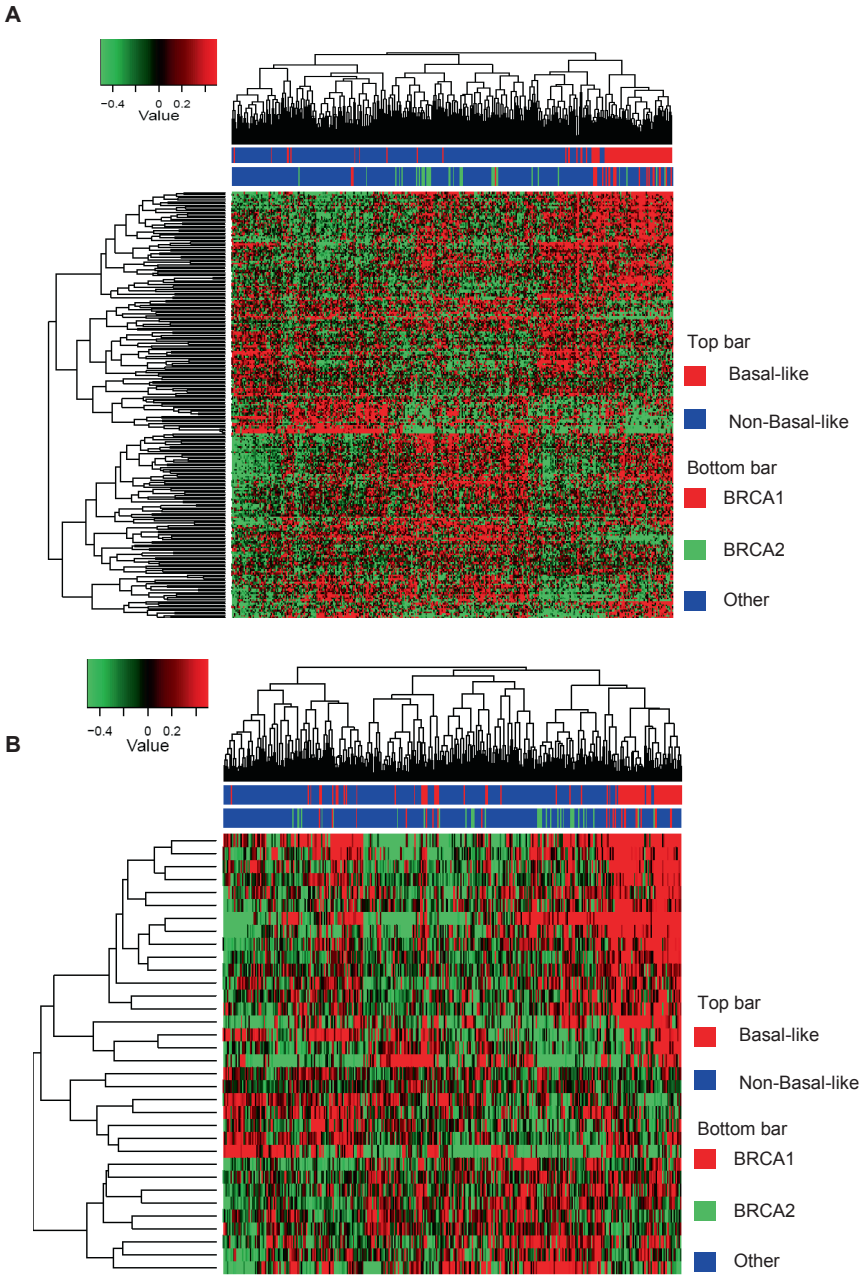


Figure 4. Hierarchical clustering of proteins mapped to gene transcripts for the Jönsson data set. **A,** Cluster analysis using the 417 up-regulated mapped proteins. The majority of the BRCA1 patients cluster together within the basal cluster in one branch of the dendrogram. The adjacent cluster contains the majority of the BRCA2 patients. **B,** Cluster analysis using the 45-protein BRCA1-deficiency signature. The BRCA1 and BRCA2 samples are now adjacent to each other in one cluster containing approximately one third of all patients.

We also assessed the performance of 1000 sets of genes randomly sampled from the whole genome and, in a more stringent approach, from the list of DNA replication, recombination and repair genes as defined by IPA (Table 5). Although both all up-regulated proteins and the BRCA1 deficiency signature achieved similar sensitivities in classifying *BRCA1/2*-mutated tumors, this was only significantly better compared to random gene lists, when using all up-regulated proteins. Nevertheless, the BRCA1 deficiency signature compared favorably to both random gene lists sampled from all genes and from random DNA repair lists, showing confidence that the classification accuracies were not obtained by chance.

In addition to leave-one-out cross-validation, we performed completely independent validation using the two other data sets containing samples with *BRCA1/2* mutation status (the combined Van de Vijver and Van 't Veer cohorts and the Waddell cohort; Table 4). The centroids constructed from the Jönsson *et al*. data set can classify BRCA1/2 samples in the Van de Vijver/Van 't Veer cohorts with a very high sensitivity of 95 % for both the up-regulated proteins and the BRCA1 deficiency signature (Table 5). A large portion of sporadic samples were assigned to the BRCA1/2 class. Because the sporadic samples were not tested for BRCA1/2 dysfunction or inactivation of other components of the homologous recombination (HR) pathway, part of these mismatching predictions reported here might reflect a true deficiency in the BRCA1/2 pathway, *i.e.*, a BRCAness phenotype[44]. For the Waddell cohort, we obtained sensitivity of 79 % and 68 % for BRCA1 patients by the up-regulated proteins and BRCA1 deficiency signature respectively (Table 5), which is comparable with the result of 74 % sensitivity reported by the authors of the data set. Our result is significant, given that the test data is completely independent from the training data, whereas internal validation was used in Waddell *et al*.[14].

These data show that the 417 up-regulated proteins in BRCA1-deficient mouse tumors, as well as the BRCA1 deficiency signature of 45 proteins can classify human BRCA1-deficient breast tumors when mapped to human transcriptomics data sets. Importantly, the classification results for the mapped mouse BRCA1 deficiency protein signature were better than the results that we obtained with the published mouse transcriptome data[24] from which we also constructed a signature using the same network-based *in silico* approach (data not shown). For example sensitivities of the protein signature for selecting BRCA1-deficient tumors were 81.8, 94.4, and 68.4 % in the Jönsson data set, the combined Van de Vijver/Van 't Veer data set and the Waddell data set, whereas these values were 63.6, 50.0 and 57.9% for the transcriptome signature (data not shown). The set of all up-regulated proteins achieved the best performance for diagnosing *BRCA1* mutations in comparison to random (DNA repair) genes. BRCA2-deficient tumors were also classified, implying enrichment for homology-directed repair-deficient tumors in general. Moreover, the 45 protein signature and all up-regulated proteins also classify a number of familial tumors without *BRCA1/2*

mutation and sporadic patients as BRCA1/2-like, suggesting that these tumors might be deficient in homology-directed DNA repair.

BRCA1 deficiency signature proteins show prognostic power when mapped to human breast cancer gene expression data sets

To investigate whether the BRCA1 deficiency proteins and signature have prognostic power, we used the mapped mRNAs of the up-regulated proteins in the four public breast cancer gene expression data sets that have associated clinical end point data (Van de Vijver, Wang, Naderi and Jönsson[8;10;12;34]; Table 4) to perform a Kaplan-Meier survival analysis. The Jönsson data set was the only cohort that has an enrichment of familial (BRCA1/2-related) patients. For comparison, we also performed a Kaplan-Meier analysis using two commercially available prognostic gene signatures (MammaPrint® and Oncotype DX®). In a third comparison, we used the Naderi signature (discovered in the Naderi cohort), which has been shown to also have prognostic power in both the Van de Vijver and Wang cohorts[10].

The mapped list of all 417 up-regulated proteins in BRCA1-deficient tumors yielded highly significant *p* values for survival analysis across all data sets, but these were only significantly better than random gene lists in the Van de Vijver data set. When sampling from a DNA repair gene background, no significant *p* values for the permutation analysis were obtained (Table 6). It is of note here that the external (commercial) gene expression-based signatures in some instances showed a similar level of underperformance when compared with random DNA repair gene lists in the sporadic data sets and were performing nonsignificantly in all permutation settings in the Jönsson cohort. Not surprisingly, the two mRNA signatures identified within their discovery cohort, (MammaPrint® in the Van de Vijver cohort[1;12], and Naderi signature in the Naderi Cohort[10]), outperformed all other signatures within their cohort.

The mapped BRCA1 deficiency signature has highly significant prognostic value. The Kaplan-Meier plots of the BRCA1 deficiency signature in the four breast cancer data sets is shown in Figure 5. Performance was comparable to the gene expression-based signatures in the three sporadic cohorts (the Van de Vijver, Wang and Caldas data sets).

Importantly, in the data set with an over-representation of familial (BRCA1/2) tumors (the Jönsson cohort), the mapped mouse BRCA1 deficiency signature outperformed all human gene expression-based signatures, and performance was still significant when compared with random (DNA-repair) gene lists. In summary, these data demonstrate that the mouse BRCA1 deficiency protein signature, when mapped to human gene expression data has prognostic value and outperforms (commercial) gene expression-based signatures in a cohort enriched for breast cancer with defects in the homology-directed DNA repair pathway.

Table 5. Performance of all 417 up-regulated proteins and the 45-protein BRCA1 deficiency signature in human gene expression data sets

True/ predicted	BRCA1	BRCA2	Familial	Sporadic	Total	Sensitivity	Specificity	All genes[a]	DNA repair background[a]
Jönsson *et al.* data set									
All 417 up-regulated proteins									
BRCA1	17	0	1	4	22	**77,3%**	84,7%	**0,016**	**0,017**
BRCA2	4	24	3	1	32	75,0%	85,9%	0,286	0,315
BRCA1/2	45	45	4	5	54	**83,3%**	69,5%	**0,040**	**0,012**
Familial	15	26	63	28	132	47,7%	78,4%		
Sporadic	32	20	45	76	173	43,9%	82,3%		
BRCA1 deficiency signature									
BRCA1	18	0	1	3	22	**81,8%**	81,3%	0,329	0,243
BRCA2	5	16	8	3	32	50,0%	86,5%	0,184	0,217
BRCA1/2	39	39	9	6	54	**72,2%**	66,6%	0,245	0,178
Familial	16	25	65	26	132	49,2%	72,7%		
Sporadic	42	19	53	59	173	34,1%	85,6%		
Combined Van de Vijver *et al.* and Van 't Veer *et al.* data sets									
All 417 up-regulated proteins									
BRCA1	17	0		1	18	**94,4%**	75,8%		
BRCA2	0	2		0	2	100,0%	57,2%		
BRCA1/2	19	19		1	20	**95,0%**	30,2%		
Sporadic	72	134		89	295	30,2%	95,0%		
BRCA1 deficiency signature									
BRCA1	17	0		1	18	**94,4%**	80,3%		
BRCA2	0	2		0	2	100,0%	71,1%		
BRCA1/2	19	19		1	20	**95,0%**	32,2%		
Sporadic	73	127		95	295	32,2%	95,0%		
Waddell *et al.* data set									
All 417 up-regulated proteins									
BRCA1	15	1	3		19	**78,9%**	78,6%		
BRCA2	7	12	11		30	40,0%	91,7%		
BRCA1/2	35	35	14		49	**71,4%**	56,0%		
Familial	8	3	14		25	56,0%	71,4%		
BRCA1 deficiency signature									
BRCA1	13	3	3		19	**68,4%**	77,5%		
BRCA2	8	8	14		30	26,7%	86,3%		
BRCA1/2	32	32	17		49	**65,3%**	52,0%		
Familial	8	4	13		25	54,0%	65,3%		

[a] *p* value of permutation analysis using random gene lists (the fraction of 1000 random gene lists of the same length performing better than the 417 up-regulated proteins or the BRCA1 deficiency signature). In the case of "All genes" sampling was done from all genes present in the human genome that had an official gene symbol. Genes from a DNA repair background were sampled from a list generated by IPA.

Poor outcome human breast tumors identified by BRCA1 deficiency signature show enrichment in p53 mutations

p53 mutations have the capacity to disrupt the signaling between accumulated DNA damage and the induction of apoptosis. Moreover, loss of functional p53 is often associated

with BRCA1-related hereditary breast cancer in humans[45;46]. For this reason, we investigated whether the poor prognosis patients identified in the survival analysis showed a significant enrichment for p53 mutations. For the Van de Vijver cohort, we were able to retrieve p53 mutational status for 204 of the 295 tumors (data not shown). Enrichment of p53 mutation in the poor prognosis patients was assessed using Fisher's exact test. Both the total list of 417 up-regulated proteins and the BRCA1 deficiency signature showed a highly significant enrichment for p53 mutations in poor prognosis patients (both p values were $< 10^{-10}$; Table 7). These data highlight the finding that the BRCA1 deficiency proteins and signature associate with p53 mutation as well as with survival.

Protein quantitation by targeted mass spectrometry

We have selected several proteins for follow-up at the protein level: four genes/proteins that showed discordant regulations: significantly up-regulated protein levels and down-regulated mRNA expression levels (NCAPD2, SIN3A, BAZ1B, TOP2B) in the BRCA1-deficient breast tumors of the mouse model. We also included one gene for which no probe was available on the microarray (TOP2A) and one protein for which protein and mRNA regulation was concordant (PARP1) in the mouse model. Of these gene products, SIN3A and TOP2B had also down-regulated mRNAs in the human data set of Jönsson, whereas PARP1 was not regulated, TOP2A was up-regulated, and for NCAPD2 and BAZ1B no probes were available. First, we confirmed the protein regulations as revealed by the spectral count data in the discovery samples using an independent measure of label-free protein quantitation, *i.e.*, the area under the curve of the extracted ion chromatograms. Second, we performed targeted mass spectrometry by SRM-MS in 10 independent mouse mammary tumors, all carcinomas.

The regulation of SIN3A, NCAPD2, TOP2A, TOP2B and PARP1 was confirmed by SRM-MS in independent tumors, with all peptides being significantly up-regulated in BRCA1-deficient breast tumors, whereas only BAZ1B was not significantly up-regulated (Supplementary Figure S3). Hierarchical clustering using all peptides from the discordant proteins clearly separated this pilot validation set according BRCA1 status (Supplementary Figure S4). In conclusion, the SRM validation of protein expression levels for which the RNA levels were discordant, underscores the fact that RNA expression levels can not always be simply translated to protein expression levels, as well as the importance of analysis of the end products of genes by proteomics.

DISCUSSION

In the present study, we aimed to identify proteins that are associated with the loss of expression of BRCA1, which is involved in homology-directed DNA repair. These proteins could potentially find use as screening, prognostic or predictive biomarkers.

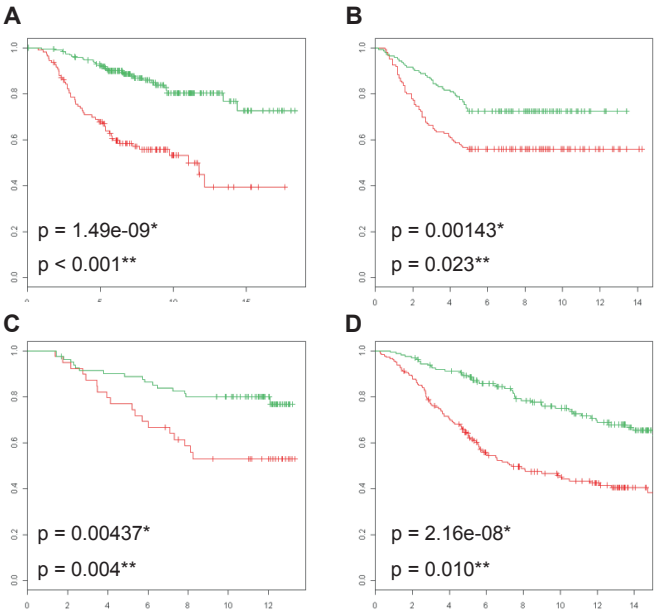


Figure 5. Survival analysis displaying Kaplan-Meier curves on a diverse set of public human gene expression breast cancer data using the BRCA1-deficiency signature. Red curves represent patients with poor prognosis and green curves are patients with good prognosis. Overall survival was used as clinical end point, unless specified otherwise. **A,** Van de Vijver data set of 295 patients. **B,** Wang cohort with 286 tumors using metastasis free survival as clinical end point. **C,** Naderi data set with 120 early stage tumors. **D,** Jönsson data set containing 359 tumors, which includes 186 familial tumors, of which 54 were confirmed BRCA1/2 carriers. * *p* value from Kaplan-Meier survival analysis**.** ** *p* value from permutation analysis (the fraction of a 1000 random gene lists of the same length performing better than the BRCA1 deficiency signature)**.**

To this end, we analyzed protein profiles in BRCA1-proficient and -deficient mouse mammary tumors using a high resolution tandem mass spectrometry-based proteomics approach. We identified 3545 proteins, of which 801 were significantly differentially regulated. A BRCA1 deficiency 45-protein signature was defined through the use of pathway and protein complex analysis, with good performance in human gene expression data sets enriched for BRCA1 deficiency. To our knowledge, this is the first comprehensive in depth proteomics analysis in genetic breast cancer. An overview of the discovery and data mining strategies is given in Figure 6.

Up-regulated proteins in BRCA1-deficient mouse mammary tumors contain basal-like markers, multiple drug targets and DNA repair(-associated) proteins

As expected, we found significant up-regulation of basal-like markers that are known to occur in breast cancer of the basal-like subtype. This is in line with the fact that human *BRCA1*-mutated tumors belong predominantly to the basal-like breast cancer subtype.

Table 6. Overview of Kaplan-Meier survival analysis in four breast cancer gene expression data sets that have associated clinical end point data

	Van de Vijver cohort				Wang cohort	
	p value[a]	# of mapped genes or proteins[b]	All genes[c]	DNA repair background[c]	p value	# of mapped genes or proteins[b]
Mouse BRCA1 deficiency proteins						
All 417 up-regulated proteins	6,12E-9	326/417	<0,001	>0,5	0,000	309/417
BRCA1 deficiency proteins	1,49E-9	43/45	<0,001	0,207	0,001	43/45
Gene expression-based signatures						
Naderi signature	1,01E-8	50/70	<0,001	0,355	6,30E-4	48/70
Oncotype DX (commercial signature)	2,22E-10	16/16	<0,001	0,029	2,07E-5	14/16
MammaPrint (commercial signature)	1,60E-14	60/61	<0,001	<0,001	0,001	47/61

[a] *p* value from Kaplan-Meier survival analysis
[b] number of proteins or genes from a signature or protein list that could be mapped to one or multiple probes on the microarray (mapped genes or proteins/all genes or proteins in the list).
[c] *p* value of permutation analysis using random (DNA repair) gene lists (the fraction of 1000 random gene lists of the same length performing better than the gene/protein list used in survival analysis). In the case of "All genes" sampling was done from all genes present in the human genome that had an official gene symbol. For this reason, genes from a DNA repair background were sampled from a list generated by IPA.

This table contains data on the Van de Vijver, Wang, Naderi and Jönsson cohorts. The top two rows represent the performance of all up-regulated BRCA1 deficiency proteins and the 45-protein BRCA1 deficiency signature mapped to matching gene symbols. The bottom three rows contain the performance of gene expression-derived signatures, as a means of comparison with the protein-based signature.

Therefore these confirmatory findings underscore the human relevance of the BRCA1-deficient mouse tumor models. BRCA1 has recently, through its function as a transcription factor, been linked to the basal transcription machinery, whereby functional BRCA1 represses transcription of basal keratins[47]. In addition, we identified a number of up-regulated proliferative markers, a feature that is more prevalent in human basal-like breast cancer.

Pathway and protein complex analysis identified DNA repair and associated processes as the most important biological function associated with the up-regulated proteins of the BRCA1-deficient tumors. This is in line with previous reports that loss of functional homology-directed DNA repair through knock-out of BRCA1 might be partially compensated for by other DNA repair mechanisms[42;48]. Importantly, we found a number of therapeutic targets to be up-regulated in the BRCA1-deficient tumors, including PARP1, TOP1, TOP2A and TOP2B.

PARP1 has been shown to be a *bona fide* drug target for human *BRCA1*-mutated tumors[43]. Up-regulation of the PARP1 protein may be a marker for the loss of functional homology-directed DNA repair in general, and might therefore be a predictive marker for the efficacy of PARP1 inhibition.

Wang cohort		Naderi cohort				Jönsson cohort			
All genes[c]	DNA repair background[c]	p value[a]	# of mapped genes or proteins[b]	All genes[c]	DNA repair background[c]	p value[a]	# of mapped genes or proteins[b]	All genes[c]	DNA repair background[c]
0,247	>0,5	0,023	262/417	0,133	0,309	9,80E-6	227/417	0,248	0,277
0,023	0,440	0,004	32/45	0,004	0,044	2,16E-8	33/45	0,010	0,008
0,008	>0,5	0,000	60/70	<0,001	0,005	1,44E-6	33/70	0,090	0,088
<0,001	0,024	0,000	13/15	<0,001	0,005	6,40E-4	13/16	0,435	0,359
0,013	>0,5	0,009	41/61	0,004	0,056	1,16E-5	35/61	0,250	0,233

Table 7. Overview and statistical analysis of enrichment for p53 mutations in poor prognosis patients versus good prognosis patients using Fisher's exact test in the Van de Vijver *et al.* data set

	Poor prognosis	Good prognosis	Sum
All 417 up-regulated proteins: *p* value = 2,20E-16[a]			
p53 mutation	55 (79%)	15 (21)	70
Wild-type p53	25 (19%)	109 (81%)	134
Sum	80	124	204
BRCA1 deficiency signature (45 proteins): *p* value = 2,55E-12[a]			
p53 mutation	49 (70%)	21 (30%)	70
Wild-type p53	26 (19%)	108 (81%)	134
Sum	75	129	204

[a] *p* value from Fisher's exact test

In line with this, the tumors of the BRCA1-deficient mice used in this study responded well to the PARP inhibitor olaparib, whereas the BRCA1-proficient mouse tumor models did not[49]. Moreover, we also found up-regulation of the topoisomerases TOP1, TOP2A and TOP2B drug targets for topotecan (TOP1 inhibitor) and doxorubicin (TOP2A and TOP2B inhibitor). These drugs inhibit the religation step of topoisomerases and therefore also induce indirectly DNA breaks. Higher levels of these proteins might have predictive value, since the BRCA1-deficient mouse tumors used in our experiments have been shown to be sensitive to topotecan[50] and doxorubicin[51]. We detected a number of other potential drug targets. HDAC1 and HDAC2, two proteins involved in chromatin remodeling by histone acetylation, were also up-regulated, although this was not significant (p = 0.07 and p = 0.14, respectively). At least 11 kinases were significantly up-regulated of which established drug targets included KIT and SRC (Supplementary Table S2, IPA drug targets). Novel kinase candidate drug targets included MAPK14, CDK9 and CSFR1. Multiple proteins

up-regulated in the BRCA1-deficient mice tumors act upstream of BRCA1 function in the homology-directed DNA repair pathway (ATM, BAZ1B, and TP53BP1), which may indicate an accumulation of these proteins in response to BRCA1 loss. In a previous study, Liu *et al.*[24] used gene expression analysis in the same mouse models as used in this study. Using gene set enrichment analysis, they reported a number of processes that were induced after BRCA1 loss, including recombinatorial repair, mitotic recombination, telomere maintenance and transcriptional regulation (e.g. chromatin remodeling).

Mouse BRCA1 deficiency protein signature with diagnostic and prognostic value in human gene expression data sets

We used an *in silico* validation approach to show that mouse proteins up-regulated in BRCA1-deficient tumors, including a BRCA1 deficiency signature, could classify human BRCA1 and BRCA2 tumors in cohorts that contained both sporadic and hereditary breast cancers. Using the BRCA1 deficiency signature, high sensitivities were achieved for classifying homology-directed DNA repair-deficient tumors in data sets known to be enriched for these patients[14]. The BRCA1 deficiency signature also classified a considerable number of sporadic and familial tumors as BRCA1/2-like. This result may be explained by the possibility that a number of sporadic and familial tumors lacking mutations in the *BRCA* genes might still harbor undetected deficiencies in homology-directed DNA repair and might therefore benefit from DNA-damaging agents. There is growing evidence that the majority of sporadic basal-like breast cancer have BRCA1 dysfunctionality other then a mutation in BRCA1 itself[9;13].

Approximately 25 % of BRCA1 tumors were not picked up by our classifier. This might be explained by the fact that our mouse BRCA1 classifier is not able to capture the full heterogeneity of all *BRCA1*-mutated breast carcinomas. In addition, a number of *BRCA1*-mutated breast carcinomas might escape detection due to restoration of homology-directed DSB repair via loss of TP53BP1[52;53] or equivalent factors.

The BRCA1 deficiency DNA repair signature showed prognostic power across a wide variety of breast cancer data sets. Moreover, our mouse protein signature outperformed two commercially available prognostic RNA-based signatures (MammaPrint® and Oncotype DX®) in a data set enriched for homology repair-deficient tumors. Finally, in breast cancer, proteins with prognostic power may have predictive value as well. Examples are the hormone receptor ESR1 and the receptor tyrosine kinase ERBB2, the expression of which predicts response to targeted therapy as well as prognosis[54].

Furthermore, patients with sporadic breast cancer identified as poor outcome by our BRCA1 deficiency signature were highly enriched for p53 mutations. Although both mouse models used to develop the BRCA1 deficiency signature were p53 deficient, this result is explained by the clinical observation that BRCA1-deficient breast cancers frequently comprise p53 mutations, and both BRCA1 and p53 alterations are enriched in triple-

negative breast cancer[45;46]. The up-regulation of drug targets involved in DNA repair (PARP1, TOP1, TOP2A and HDAC1), may indicate the predictive potential of our BRCA1 deficiency DNA repair candidates. We were not able to verify the predictive potential of our BRCA1 deficiency protein signature in large cohorts of treated breast cancers, since the therapeutic agents (PARP1 and TOP inhibition, cisplatin treatment, HDAC inhibition) are still in clinical trial phase, so no large-scale publicly available gene-expression data sets exist to date.

Several breast cancer proteomic studies have been reported to date. Biological materials used ranged from mouse tissue[21;55], human breast cancer cell lines and tissues[18;20;21;56-58]. A few studies yielded a number of markers with potential for treatment prediction. Umar *et al*.[19] have recently identified a protein profile in microdissected breast tumor cells putatively predictive for the efficacy of tamoxifen. Moreover, the differential up-regulation and activity of a number of kinases across a panel of breast cancer cell lines correlated with differential responsiveness to small molecule inhibitors in these cancer cell lines[59].

Concluding remarks

Our study demonstrates that in-depth high resolution proteomics of tumor tissue from different mouse models is a successful strategy to discover candidate protein biomarkers with screening, prognostic and possibly also predictive potential for human BRCA1 and homology-directed double-strand break repair-deficient breast tumors. The proteins up-regulated in mammary tumors from mouse models with a deficiency in BRCA1 are enriched in DNA repair(-associated) proteins, which points towards a potential rescue mechanism for the loss of homology-directed DNA repair. In addition, a pathway in conjunction with protein complex analysis has proven to be a promising strategy to construct a signature that has diagnostic and prognostic potential across multiple human breast cancer gene expression data sets. This signature shows specificity for BRCA1 and homology-directed DNA repair deficiency and has high prognostic potential in breast cancer data sets enriched with homology-directed repair-deficient tumors. Several up-regulated DNA repair proteins within this signature have been shown to be drug targets in homology-directed DNA repair-deficient tumors, suggesting that they may have predictive power for tailored therapies. Because multiple drug targets are up-regulated, these tumors might also benefit from combination therapy.

Finally, we point out that the BRCA1 deficiency transcriptome signature that we obtained by mapping mouse BRCA1 deficiency-associated breast tumor proteins is novel and could not be obtained by using the published mouse transcriptome data[24] as a starting point. To date, there is only one BRCAness gene expression signature reported for ovarian cancer[60]. However, this signature was developed using a publicly available ovarium cancer transcriptomics data set and with a pilot study for predictiveness based on only 10 BRCA mutated/reverted samples originating from 6 patients and this signature was not externally

evaluated in multiple large (BRCA1/2-deficient) breast cancer data sets. Together, these results underscore the novelty of our BRCAness transcriptome signature that we obtained by mapping mouse BRCA1 deficiency-associated breast tumor proteins.

Future studies should address the value of our BRCA1 deficiency signature both at the transcriptome and proteome level for patient selection for treatment in breast cancer and other tumors types with potential homology repair deficiencies. With the advent of targeted mass spectrometry methods like SRM-MS, the signature proteins may be analyzed in pretreatment biopsies in one multiplex analysis, without the need for antibodies. Targeted multiplex analysis in aspirate fluid and plasma may highlight their potential use for non-invasive testing.

METHODS

Materials

All chemicals, unless otherwise specified, were obtained from Sigma (Sigma Aldrich, Zwijndrecht, The Netherlands). HPLC solvents, LC-MS grade water, acetonitrile and formic acid, were obtained from Biosolve (Biosolve B.V., Valkenswaard, The Netherlands). Porcine sequence-grade modified trypsin was obtained from Promega (Promega Benelux B.V., Leiden, The Netherlands).

Mouse strains and tumors

Generation of conditional mutants and K14*cre* transgenic mice has been described previously[24;25]. All animal experiments were approved by the Animal Ethics Committee of the Netherlands Cancer Institute (NKI). When grown to a size of approximately 500 mm^3, tumors were dissected, snap frozen and stored at -80 °C until use.

Tissue homogenization and fractionation using gel electrophoresis

For homogenization, we cut a piece of ~20 mg in a bath of liquid nitrogen in smaller parts. The proteins in the mammary tumors tissue samples were solubilised in 800 µL 1x reducing sodium dodecyl sulfate (SDS) sample buffer (containing 62.5 mM Tris-HCl, 2 % w/v SDS, 10 % v/v glycerol, and 0.0025 % bromophenol blue, 100 mM DTT, pH 6.8) using a Pellet Pestles microgrinder system (Kontes glassware, Vineland, NJ). Subsequently the proteins were denatured by heating at 100 °C for 10 min. Any insoluble debris was removed by centrifuging for 15 min at maximum speed (16.1 rcf) in a benchtop centrifuge.

The proteins were fractionated using one-dimensional SDS-polyacrylamide gel electrophoresis (SDS-PAGE). 25 µL of each homogenized sample (containing about 50 µg protein) was loaded on a well of a pre-cast NuPAGE 4–12 % w/v Bis-Tris 1.5-mm minigel (Invitrogen). The stacking gel contained 4 % w/v acrylamide/Bis-Tris. Electrophoresis was

carried out at 200 V in NuPAGE MES SDS running buffer (50 mM Tris base, 50 mM MES, 0.1 % w/v SDS, 1 mM EDTA, pH 7.3) until the dye front reached the end of the gel. Following electrophoresis, gels were fixed with a solution of 50 % ethanol and 3 % phosphoric acid. Staining was carried out in a solution of 34 % methanol, 3 % phosphoric acid, 15 % ammonium sulfate and 0.1 % Coomassie Blue G-250 (Bio-Rad) with subsequent destaining in MilliQ water.

In-gel digestion and nanoLC-FTMS

For in-gel digestion, gel lanes were cut in 10 bands and each band was processed according to the method of Shevchenko *et al.*[26]. Briefly, the bands were washed and dehydrated three times in 50 mM ammonium bicarbonate (ABC, pH 7.9) / 50 mM ABC + 50% acetonitrile (ACN). Subsequently, cysteine bonds were reduced with 10 mM DTT (dithiotreitol) for 1 h at 56°C and alkylated with 50 mM iodoacetamide for 45 minutes at room temperature in the dark. After two subsequent wash/dehydration cycles the bands were dried 10 min in a vacuum centrifuge and incubated overnight with 0.06 µg/µl trypsin at 25 °C. Peptides were extracted once in 1% formic acid and subsequently twice in 50% ACN in 5% formic acid. The volume was reduced to 50 µl in a vacuum centrifuge prior to LC-MS analysis.

Peptides were separated by an Ultimate 3000 nanoLC system (Dionex LC-Packings, Amsterdam, The Netherlands) equipped with a 20-cm x 75-µm inner diameter fused silica column custom packed with 3-µm 100 Å ReproSil Pur C18 aqua (Dr Maisch GMBH, Ammerbuch-Entringen, Germany) as described before[27]. After injection, the peptides were trapped at 30 µl/min on a 0.5-cm x 300-µm inner diameter Pepmap C18 cartridge (Dionex LC-Packings, Amsterdam, The Netherlands) at 2% buffer B (buffer A: 0.05% formic acid in MQ; buffer B: 80 % ACN + 0.05% formic acid in MQ) and separated at 300 nl/min in a 10-40 % buffer B gradient in 60 min. Eluting peptides were ionized at 1.7 kV in a Nanomate Triversa Chip-based nanospray source using a Triversa LC coupler (Advion, Ithaca, NJ). Intact peptide mass spectra and fragmentation spectra were acquired on a LTQ-FT hybrid mass spectrometer (Thermo Fisher, Bremen, Germany). Intact masses were measured at resolution 50.000 in the ICR cell using a target value of 1×10^6 charges. In parallel, following an FT prescan, the top five peptide signals (charge states 2+ and higher) were submitted to MS/MS in the linear ion trap (3-atomic mass unit isolation width, 30-ms activation, 35 % normalized activation energy, Q value of 0.25, and a threshold of 5000 counts). Dynamic exclusion was applied with a repeat count of 1 and an exclusion time of 30 s.

LC-SRM-analyses

Independent BRCA1-deficient and -proficient mouse breast tumors (n = 5 in each group) were analysed in triplicate on an Ultimate 3000 RSCL Nanosystem (Dionex) that was hyphenated to an QTRAP® 5500 instrument (AB SCIEX, Foster City, CA) operated in positive

SRM mode and equipped with a nano-electrospray source with applied voltage of 2.404 kV and a capillary heater temperature of 225 °C. The Nanoflow LC system and QTRAP® 5500 system were both controlled using Analyst 1.5.1 Software. The combined information from each SRM information dependent acquisition (IDA) experiment was used to perform Mascot searches against the international protein index (IPI) mouse database v3.65 and MultiQuant™ software version 2.1 (AB SCIEX).

The scheduled SRM mode comprised the following parameters: SRM detection window of 420 s, target scan time of 3.0 s, curtain gas of 15, ion source gas 1 of 15, declustering potential of 80, and entrance potential of 10. Q1 resolution was set to unit and Q3 resolution was set to unit. Pause between mass ranges was set to 1 ms. Collision cell exit potentials (CXP)was set to 36 for all transitions. Peak integration was performed using MultiQuant™ software version 2.1 (AB SCIEX) software and manually reviewed.

Chromatographic separation of peptides was performed by a 68-min gradient at 300 nl/min. Solvent A (0.05 % formic acid water) and solvent B (0.05 % formic acid, 80% acetonitrile) were mixed at 2 % B from 0 to 3 min, 15 % B at 4 min, 36 % B at 49 min, 99 % B from 50 to 54 min, and 2 % B at 55 to 68 min. The nano-LC columns were made in house and consisted of 20-cm x 75-µm inner diameter fused silica custom packed with 3-µm 100 Å ReproSil Pur C18 aqua (Dr Maisch GMBH, Ammerbuch-Entringen, Germany) as described before[27]. After injection, peptides were trapped at 6 µl/min at 2 % buffer B.

An SRM assay for the target proteins (NCAPD2, SIN3A, BAZ1B, TOP2A, TOP2B, PARP1) was developed using the MRMPilot™ software version 2.1 from AB SCIEX. The software requires an amino acid sequence of the protein of interest, a starter method containing the LC conditions, and an empty SRM-IDA experiment. The software performs an *in silico* digest of the protein and creates a set of peptides that would result after full tryptic digestion. For each of these peptides, it will generate an SRM transition for the calculated m/z of the precursor ion and an appropriate fragment ion. Assay development subsequently entails verification of the peptides and CE optimisation of the transitions, both in multiplexed LC-SRM analyses. During verification, the highest responding peptides/transitions at a theoretically calculated optimum CE energy are determined, as well as the identity of the peptide via SRM triggered MS/MS. During CE optimisation the transitions selected after verification are optimised during the chromatographic elution of the peptide.

For verification, a mixture of samples previously analysed using FTMS and indicating abundance of the target candidates was analysed in 10 unscheduled SRM analyses to find the highest responding tryptic peptides from the target proteins, as well as their elution time during the chromatographic run. For each peptide 10 theoretically predicted transitions were assessed for detection response and identity. Identity was confirmed using MIDAS (MRM Initiated Detection and Sequencing) with a threshold of 500 counts for an SRM transition response to trigger two MS/MS spectra of the peptide to be acquired at

rolling collision energy. Each of the 10 verification analyses was set up to detect 289 of all theoretically predicted transitions and their theoretically predicted optimum collision energy for all theoretically predicted peptides that can result after tryptic digestion of the candidate proteins. The total scan time for each cycle of the instrument during verification was 3.757 s, resulting in a dwell time of 10 ms for each transition in the unscheduled verification analyses.

For CD optimization, all data of unscheduled analyses were uploaded to the MRMPilot, which was set to select the five best detected transitions for each peptide and assign a chromatographic retention time to each peptide. Subsequently collision energy for each transition was optimised in 13 LC-SRM analyses, with each analysis set-up to detect 104 scheduled transitions that resulted from verification, at nine different collision energies, centred at 3 V intervals around the theoretically predicted optimum with a dwell time of 25 ms. All data of CE optimisation cycles were uploaded to the MRMPilot and for each peptide three transitions at the experimentally found optimum and the experimentally found retention time were included in the final assay. The final assay contained 129 scheduled transitions, three for each peptide, with one to five peptides for each of the seven candidate proteins.

Data analysis

Protein identification – MS/MS spectra were searched against the mouse IPI database (56.555 entries) using Sequest (version 27, revision 12), which is part of the BioWorks 3.3 data analysis package (Thermo Fisher, San Jose, CA). MS/MS spectra were searched with a maximum allowed deviation of 10 ppm for the precursor mass and 1 atomic mass unit for fragment masses. Methionine oxidation and cysteine caboxamidomethylation were allowed modifications, two missed cleavages were allowed and the minimum number of tryptic termini was 1. After database searching, the DTA and OUT files were imported into Scaffold 1.07 (Proteome software, Portland, OR). Scaffold was used to organize the gel-band data and to validate peptide identifications using the Peptide Prophet algorithm[28;29]. Only identifications with a probability of >95 % were retained. Subsequently, the Protein Prophet algorithm was applied and protein identifications with a probability of >99 % with two peptides or more in at least sample were retained. The false discovery rate for the detected proteins using this workflow is on average around 0.5 %, and was not calculated again. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony. For each protein identified, the total number of MS/MS spectra detected for each protein identified (spectral counts) was exported to Excel 2003 (Microsoft, Redmond, USA).

Spectral count normalization and statistics – Normalization was performed as described previously[30;31]. The spectral counts of each protein were divided by the total spectral counts

of all proteins within a sample. This number was multiplied with a constant equal to the average of total spectral counts of all samples to obtain a normalized spectral count value in the same range as the non-normalized spectral counts. The beta-binomial test[30] was applied to find proteins that show significant differences in spectral count numbers between the tumor group and the reference group. Proteins with a *p* value less than 0.05 were designated as being significant. Hierarchical clustering was carried out using R. For analysis of reproducibility, we calculated the average coefficient of variation (CV) of the normalized spectral counts from overlapping proteins for three technical replicates.

SRM data analysis – Technical replicates were removed until CV of all triplicate analysis was <20 %. Subsequently, in each remaining analysis, the ratio of the AUC of Transition1/Transition2, Transition 2/Transition3 and Transition1/Transition3 was calculated. The two transitions resulting in the lowest CV percentage over all analyses were selected for further calculations; the sum of the AUC of these two transitions was determined in each sample, and a fold change for each peptide between the groups was determined by the ratio of the summed AUC in each group. The average of the fold changes of peptides belonging to one protein was determined for each protein. When the CV percentage of the average of the fold changes of the peptides of one protein was >10 %, the transitions of these peptides were visually inspected and excluded when co-eluting false positive responses were observed that had not been detected by Multiquant smoothing and peak splitting algorithms or in-house developed R-script processing. The calculated levels for each approved peptide were normalised on the level of Tuba1b in each sample.

Pathway analysis – The list of identified proteins was uploaded into the Ingenuity Pathways Analysis (IPA) software (Ingenuity Systems, Redwood City, CA) as a tab-delimited text file containing IPI accession numbers, *p* values, and fold changes calculated with a correction factor (adding 0.5 to the spectral counts of all proteins before normalization). The proteins were uploaded and mapped to the corresponding "gene objects" in the Ingenuity Pathways knowledge base. Functional analysis was performed to identify the high level biological functions that were most significantly associated to the differentially regulated proteins in the data set. Significantly regulated proteins within the high level functions are displayed graphically as nodes (proteins/gene objects) and edges (the biological relationships between the nodes). All edges are supported by at least one reference from the literature, textbook or canonical information stored in the Ingenuity knowledge base. Ingenuity Pathways Analysis computes one or more *p* values for each specific function within a high level function according to the fit of the user's set of significant proteins. The significance of functional enrichment is computed by a Fisher's exact test. Finally, the Path Designer feature was used to create graphically rich network images. In addition, we used the COFECO tool for the mapping of significantly differentially regulated proteins to protein complexes[32]. The obtained complexes were further visualized using STRING[33] and Cytoscape, respectively.

Human gene expression data sets

To explore the diagnostic and prognostic value of the protein expression data from the mouse models, we made use of publicly available human gene expression data sets. To map the up-regulated mouse BRCA1 deficiency proteins to public data sets of human arrays, we first matched mouse gene symbols to human gene symbols using the BioMart website (http://www.biomart.org). We used layout documentation files for the various microarray platforms from Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo), MIAMExpress (http://www.ebi.ac.uk/miamexpress/) or Rosetta Inpharmatics (http://www.rii.com/publications/default.html/) to retrieve the matching gene symbols on each platform. The following human breast cancer data sets were used: (i) Van de Vijver data set[12]. A validation study of a prognostic gene expression signature (MammaPrint®), which included 295 young patients with early stage breast cancer, of which 151 were lymph node negative, 226 were estrogen receptor-positive, and 110 had received adjuvant chemotherapy. We were also able to retrieve p53 mutational status for 204 tumors in this data set (data not shown). (ii) Van't Veer data set[1]. In this discovery study for a prognostic signature (MammaPrint®), the authors also analyzed 18 BRCA1 and 2 BRCA2 samples on the same platform used for the Van de Vijver data set[12]. (iii) E-UCON-1 data set [10] (subsequently referred to as the Naderi data set). This data set was used for discovery of a prognosis profile in a set of women with early stage breast cancer representative of breast cancer demographics. Of the 132 breast cancer tissues, we used a subset of 120 patients for survival analysis that had the same orientation in dye labeling concerning the reference and tumor samples and that also had associated survival data. (iv) GSE2034 data set[34] (subsequently referred to as Wang data set). This was a discovery and validation analysis of a gene signature for the prediction of breast cancer patient outcomes. It consists of 286 lymph node-negative breast cancer patients who never received adjuvant chemotherapy and of which 209 were estrogen receptor-positive. We logged the normalized intensity values and performed zero mean and unit variance normalization. (v) GSE22133 data set[8] (subsequently referred to as the Jönsson data set). This discovery data set consists of 359 breast tumors including 186 familial, of which 22 were BRCA1-mutated and 32 were BRCA2 mutated. (vi) GSE19177 data set[14] (subsequently referred to as the Waddell data set). This data set contains familial tumors only. Nineteen had a BRCA1 mutation, 30 had a BRCA2 mutation, whereas 25 did not have an identifiable mutation. One tumor was excluded from analysis because it had unknown mutational status. For all data sets, we used the normalized log ratios in the analyses, unless specified otherwise above.

Centroid classification and survival analysis

We used a nearest centroid classifier to test the diagnostic and prognostic power of the mapped protein/gene signature on the public human gene expression data sets in

combination with leave-one-out-cross-validation. First, the signature proteins/genes in the validation sets were identified. We used a centroid classification scheme to assess BRCA1 and homology-directed DNA repair deficiency, whereby centroids were built by taking the average expression value for each signature gene in the diagnostic groups, excluding the leave-out sample. The leave-out samples were then classified into different diagnostic groups using the nearest correlation criterion. For classification with a centroid on external data sets, genes were collapsed by taking the median across all probes. This centroid classification scheme was also used for classifications in the Kaplan-Meier survival analysis. In all data sets, patients who survived 5 years or more constituted the good prognosis group (centroid), while patients who survived less then 5 years were used for the poor prognosis group (centroid)[10;12;34]. The average expression value for each signature gene in the good and poor prognosis centroid was computed without the leave-out sample. The leave-out samples were then classified into good or poor prognostic groups using the nearest correlation criterion. To see if a gene list performed better than random, both in the diagnostic and in the survival analysis, we also ran analysis with 1000 random gene lists of the same size using the same scheme. We only included probes on the arrays which were annotated with a gene symbol. The same scheme was applied for the prognostics mRNA based signatures used as a comparison.

ACKNOWLEDGMENTS

We would like to acknowledge Anne-Lise Borresen, Anita Langerod and Hugo Horlings for generation and access to p53 mutation data from the Van de Vijver cohort and Davide Chiasserini for generating the protein complex images using Cytoscape. We also acknowledge Piet Borst for careful reading and adjusting this manuscript. This research was supported by the CenE/Van Lanschot (MW) and the VUmc-Cancer Center Amsterdam (CRJ, TVP and proteomics infrastructure), the Dutch Cancer Society (project grants to SR and JJ) and the Netherlands Organization for Scientific Research (NWO-Vidi grant to SR and NWO Cancer Systems Biology Center (CSBC) grant to JJ). JEJ was supported by a Toptalent fellowship from the Netherlands Organization for Scientific Research (NWO).

REFERENCES

1. Van ‘t Veer, L. J., Dai, H., van, d., V, He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der, K. K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R., and Friend, S. H. (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415, 530-536.
2. Turner, N. C., Reis-Filho, J. S. (2006) Basal-like breast cancer and the BRCA1 phenotype. *Oncogene* 25, 5846-5853.

3. Jaspers, J. E., Rottenberg, S., and Jonkers, J. (2009) Therapeutic options for triple-negative breast cancers with defective homologous recombination. *Biochim. Biophys. Acta* 1796, 266-280.
4. Gudmundsdottir, K., Ashworth, A. (2006) The roles of BRCA1 and BRCA2 and associated proteins in the maintenance of genomic stability. *Oncogene* 25, 5864-5874.
5. Tutt A., Robson M., Garber J.E., Domchek S.M., Audeh M.W., Weitzel J.N., Friedlander M., Arun B., Loman N., Schmutzler R.K., Wardley A., Mitchell G., Earl H., Wickens M., Carmichael J. (2010) Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and advanced breast cancer: a proof-of-concept trial. *Lancet* 376, 235-244.
6. Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de, R. M., Jeffrey, S. S., Thorsen, T., Quist, H., Matese, J. C., Brown, P. O., Botstein, D., Eystein, L. P., and Borresen-Dale, A. L. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA* 98, 10869-10874.
7. Glas, A. M., Floore, A., Delahaye, L. J., Witteveen, A. T., Pover, R. C., Bakx, N., Lahti-Domenici, J. S., Bruinsma, T. J., Warmoes, M. O., Bernards, R., Wessels, L. F., and Van't Veer, L. J. (2006) Converting a breast cancer microarray signature into a high-throughput diagnostic test. *BMC Genomics* 7, 278.
8. Jonsson, G., Staaf, J., Vallon-Christersson, J., Ringner, M., Holm, K., Hegardt, C., Gunnarsson, H., Fagerholm, R., Strand, C., Agnarsson, B. A., Kilpivaara, O., Luts, L., Heikkila, P., Aittomaki, K., Blomqvist, C., Loman, N., Malmstrom, P., Olsson, H., Johannsson, O. T., Arason, A., Nevanlinna, H., Barkardottir, R. B., and Borg, A. (2010) Genomic subtypes of breast cancer identified by array-comparative genomic hybridization display distinct molecular and clinical characteristics. *Breast Cancer Res.* 12, R42.
9. Joosse, S. A., Brandwijk, K. I., Mulder, L., Wesseling, J., Hannemann, J., and Nederlof, P. M. (2011) Genomic signature of BRCA1 deficiency in sporadic basal-like breast tumors. *Genes Chromosomes Cancer* 50, 71-81.
10. Naderi, A., Teschendorff, A. E., Barbosa-Morais, N. L., Pinder, S. E., Green, A. R., Powe, D. G., Robertson, J. F., Aparicio, S., Ellis, I. O., Brenton, J. D., and Caldas, C. (2007) A gene-expression signature to predict survival in breast cancer across independent data sets. *Oncogene* 26, 1507-1516.
11. Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F. L., Walker, M. G., Watson, D., Park, T., Hiller, W., Fisher, E. R., Wickerham, D. L., Bryant, J., and Wolmark, N. (2004) A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* 351, 2817-2826.
12. van de Vijver, M., He, Y. D., Van 't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., van, d., V, Bartelink, H., Rodenhuis, S., Rutgers, E. T., Friend, S. H., and Bernards, R. (2002) A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* 347, 1999-2009.
13. Vollebergh, M. A., Lips, E. H., Nederlof, P. M., Wessels, L. F., Schmidt, M. K., van Beers, E. H., Cornelissen, S., Holtkamp, M., Froklage, F. E., de Vries, E. G., Schrama, J. G., Wesseling, J., van, d., V, van, T. H., de, B. M., Hauptmann, M., Rodenhuis, S., and Linn, S. C. (2010) An aCGH classifier derived from BRCA1-mutated breast cancer and benefit of high-dose platinum-based chemotherapy in HER2-negative breast cancer patients. *Ann. Oncol.* Advance Access.
14. Waddell, N., Arnold, J., Cocciardi, S., da, S. L., Marsh, A., Riley, J., Johnstone, C. N., Orloff, M., Assie, G., Eng, C., Reid, L., Keith, P., Yan, M., Fox, S., Devilee, P., Godwin, A. K., Hogervorst, F. B., Couch, F., Grimmond, S., Flanagan, J. M., Khanna, K., Simpson, P. T., Lakhani, S. R., and Chenevix-Trench, G. (2010) Subtypes of familial breast tumours revealed by expression and copy number profiling. *Breast Cancer Res. Treat.* 123, 661-677.
15. Straver, M. E., Glas, A. M., Hannemann, J., Wesseling, J., van, d., V, Rutgers, E. J., Vrancken Peeters, M. J., van, T. H., Van't Veer, L. J., and Rodenhuis, S. (2010) The 70-gene signature as a response predictor for neoadjuvant chemotherapy in breast cancer. *Breast Cancer Res. Treat.* 119, 551-558.
16. Cardoso, F., Piccart-Gebhart, M., Van't, V. L., and Rutgers, E. (2007) The MINDACT trial: the first prospective clinical validation of a genomic tool. *Mol. Oncol.* 1, 246-251.

17. Pavlou, M. P., Kulasingam, V., Sauter, E. R., Kliethermes, B., and Diamandis, E. P. (2010) Nipple aspirate fluid proteome of healthy females and patients with breast cancer. *Clin. Chem.* 56, 848-855.
18. Xu, X., Qiao, M., Zhang, Y., Jiang, Y., Wei, P., Yao, J., Gu, B., Wang, Y., Lu, J., Wang, Z., Tang, Z., Sun, Y., Wu, W., and Shi, Q. (2010) Quantitative proteomics study of breast cancer cell lines isolated from a single patient: discovery of TIMM17A as a marker for breast cancer. *Proteomics* 10, 1374-1390.
19. Umar, A., Kang, H., Timmermans, A. M., Look, M. P., Meijer-van Gelder, M. E., den Bakker, M. A., Jaitly, N., Martens, J. W., Luider, T. M., Foekens, J. A., and Pasa-Tolic, L. (2009) Identification of a putative protein profile associated with tamoxifen therapy resistance in breast cancer. *Mol. Cell Proteomics* 8, 1278-1294.
20. Lai, T. C., Chou, H. C., Chen, Y. W., Lee, T. R., Chan, H. T., Shen, H. H., Lee, W. T., Lin, S. T., Lu, Y. C., Wu, C. L., and Chan, H. L. (2010) Secretomic and proteomic analysis of potential breast cancer markers by two-dimensional differential gel electrophoresis. *J. Proteome Res.* 9, 1302-1322.
21. Celis, J. E., Gromov, P., Cabezon, T., Moreira, J. M., Ambartsumian, N., Sandelin, K., Rank, F., and Gromova, I. (2004) Proteomic characterization of the interstitial fluid perfusing the breast tumor microenvironment: a novel resource for biomarker and therapeutic target discovery. *Mol. Cell Proteomics* 3, 327-344.
22. Geiger, T., Cox, J., Ostasiewicz, P., Wisniewski, J. R., and Mann, M. (2010) Super-SILAC mix for quantitative proteomics of human tumor tissue. *Nat. Methods* 7, 383-385.
23. Becker, S., Cazares, L. H., Watson, P., Lynch, H., Semmes, O. J., Drake, R. R., and Laronga, C. (2004) Surfaced-enhanced laser desorption/ionization time-of-flight (SELDI-TOF) differentiation of serum protein profiles of BRCA-1 and sporadic breast cancer. *Ann. Surg. Oncol.* 11, 907-914.
24. Liu, X., Holstege, H., van der, G. H., Treur-Mulder, M., Zevenhoven, J., Velds, A., Kerkhoven, R. M., van Vliet, M. H., Wessels, L. F., Peterse, J. L., Berns, A., and Jonkers, J. (2007) Somatic loss of BRCA1 and p53 in mice induces mammary tumors with features of human BRCA1-mutated basal-like breast cancer. *Proc. Natl. Acad. Sci. USA* 104, 12111-12116.
25. Derksen, P. W., Liu, X., Saridin, F., van der, G. H., Zevenhoven, J., Evers, B., van, B., Jr., Griffioen, A. W., Vink, J., Krimpenfort, P., Peterse, J. L., Cardiff, R. D., Berns, A., and Jonkers, J. (2006) Somatic inactivation of E-cadherin and p53 in mice leads to metastatic lobular mammary carcinoma through induction of anoikis resistance and angiogenesis. *Cancer Cell* 10, 437-449.
26. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* 68, 850-858.
27. Piersma, S. R., Fiedler, U., Span, S., Lingnau, A., Pham, T. V., Hoffmann, S., Kubbutat, M. H., and Jimenez, C. R. (2010) Workflow comparison for label-free, quantitative secretome proteomics for cancer biomarker discovery: method evaluation, differential analysis, and verification in serum. *J. Proteome Res.* 9, 1913-1922.
28. Nesvizhskii, A. I., Keller, A., Kolker, E., and Aebersold, R. (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* 75, 4646-4658.
29. Keller, A., Nesvizhskii, A. I., Kolker, E., and Aebersold, R. (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* 74, 5383-5392.
30. Pham, T. V., Piersma, S. R., Warmoes, M., and Jimenez, C. R. (2010) On the beta-binomial model for analysis of spectral count data in label-free tandem mass spectrometry-based proteomics. *Bioinformatics* 26, 363-369.
31. Albrethsen, J., Knol, J. C., Piersma, S. R., Pham, T. V., de, W. M., Mongera, S., Carvalho, B., Verheul, H. M., Fijneman, R. J., Meijer, G. A., and Jimenez, C. R. (2010) Subnuclear proteomics in colorectal cancer: identification of proteins enriched in the nuclear matrix fraction and regulation in adenoma to carcinoma progression. *Mol. Cell Proteomics* 9, 988-1005.
32. Sun, C. H., Kim, M. S., Han, Y., and Yi, G. S. (2009) COFECO: composite function annotation enriched by protein complex data. *Nucleic Acids Res.* 37, W350-W355.
33. Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., Bork, P., and von, M. C. (2009) STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* 37, D412-D416.

34. Wang, Y., Klijn, J. G., Zhang, Y., Sieuwerts, A. M., Look, M. P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M. E., Yu, J., Jatkoe, T., Berns, E. M., Atkins, D., and Foekens, J. A. (2005) Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 365, 671-679.
35. Wright, M. H., Calcagno, A. M., Salcido, C. D., Carlson, M. D., Ambudkar, S. V., and Varticovski, L. (2008) Brca1 breast tumors contain distinct CD44+/. *Breast Cancer Res.* 10, R10.
36. Stuart-Harris, R., Caldas, C., Pinder, S. E., and Pharoah, P. (2008) Proliferation markers and survival in early breast cancer: a systematic review and meta-analysis of 85 studies in 32,825 patients. *Breast* 17, 323-334.
37. Cordes, N. (2006) Integrin-mediated cell-matrix interactions for prosurvival and antiapoptotic signaling after genotoxic injury. *Cancer Lett.* 242, 11-19.
38. Wang, Y., Cortez, D., Yazdi, P., Neff, N., Elledge, S. J., and Qin, J. (2000) BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. *Genes Dev.* 14, 927-939.
39. Heale, J. T., Ball, A. R., Jr., Schmiesing, J. A., Kim, J. S., Kong, X., Zhou, S., Hudson, D. F., Earnshaw, W. C., and Yokomori, K. (2006) Condensin I interacts with the PARP-1-XRCC1 complex and functions in DNA single-strand break repair. *Mol. Cell* 21, 837-848.
40. Lee, C. G., Hague, L. K., Li, H., and Donnelly, R. (2004) Identification of toposome, a novel multisubunit complex containing topoisomerase IIalpha. *Cell Cycle* 3, 638-647.
41. Osley, M. A., Tsukuda, T., and Nickoloff, J. A. (2007) ATP-dependent chromatin remodeling factors and DNA damage repair. *Mutat. Res.* 618, 65-80.
42. Gudmundsdottir, K., Ashworth, A. (2006) The roles of BRCA1 and BRCA2 and associated proteins in the maintenance of genomic stability. *Oncogene* 25, 5864-5874.
43. Fong, P. C., Boss, D. S., Yap, T. A., Tutt, A., Wu, P., Mergui-Roelvink, M., Mortimer, P., Swaisland, H., Lau, A., O'Connor, M. J., Ashworth, A., Carmichael, J., Kaye, S. B., Schellens, J. H., and de Bono, J. S. (2009) Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. *N. Engl. J. Med.* 361, 123-134.
44. Turner, N., Tutt, A., and Ashworth, A. (2004) Hallmarks of 'BRCAness' in sporadic cancers. *Nat. Rev. Cancer* 4, 814-819.
45. Manie, E., Vincent-Salomon, A., Lehmann-Che, J., Pierron, G., Turpin, E., Warcoin, M., Gruel, N., Lebigot, I., Sastre-Garau, X., Lidereau, R., Remenieras, A., Feunteun, J., Delattre, O., de, T. H., Stoppa-Lyonnet, D., and Stern, M. H. (2009) High frequency of TP53 mutation in BRCA1 and sporadic basal-like carcinomas but not in BRCA1 luminal breast tumors. *Cancer Res.* 69, 663-671.
46. Holstege, H., Joosse, S. A., van Oostrom, C. T., Nederlof, P. M., de, V. A., and Jonkers, J. (2009) High incidence of protein-truncating TP53 mutations in BRCA1-related breast cancer. *Cancer Res.* 69, 3625-3633.
47. Gorski, J. J., James, C. R., Quinn, J. E., Stewart, G. E., Staunton, K. C., Buckley, N. E., McDyer, F. A., Kennedy, R. D., Wilson, R. H., Mullan, P. B., and Harkin, D. P. (2010) BRCA1 transcriptionally regulates genes associated with the basal-like phenotype in breast cancer. *Breast Cancer Res. Treat.* 122, 721-731.
48. Helleday, T., Petermann, E., Lundin, C., Hodgson, B., and Sharma, R. A. (2008) DNA repair pathways as targets for cancer therapy. *Nat. Rev. Cancer* 8, 193-204.
49. Rottenberg, S., Jaspers, J. E., Kersbergen, A., van der, B. E., Nygren, A. O., Zander, S. A., Derksen, P. W., de, B. M., Zevenhoven, J., Lau, A., Boulter, R., Cranston, A., O'Connor, M. J., Martin, N. M., Borst, P., and Jonkers, J. (2008) High sensitivity of BRCA1-deficient mammary tumors to the PARP inhibitor AZD2281 alone and in combination with platinum drugs. *Proc. Natl. Acad. Sci. USA* 105, 17079-17084.
50. Zander, S. A., Kersbergen, A., van der, B. E., de, W. N., van, T. O., Gunnarsdottir, S., Jaspers, J. E., Pajic, M., Nygren, A. O., Jonkers, J., Borst, P., and Rottenberg, S. (2010) Sensitivity and acquired resistance of BRCA1;p53-deficient mouse mammary tumors to the topoisomerase I inhibitor topotecan. *Cancer Res.* 70, 1700-1710.
51. Rottenberg, S., Nygren, A. O., Pajic, M., van Leeuwen, F. W., van, d. H., I, van de, W. K., Liu, X., de Visser, K. E., Gilhuijs, K. G., van, T. O., Schouten, J. P., Jonkers, J., and Borst, P. (2007) Selective induction of chemotherapy resistance of mammary tumors in a conditional mouse model for hereditary breast cancer. *Proc. Natl. Acad. Sci. USA* 104, 12117-12122.

52. Bouwman, P., Aly, A., Escandell, J. M., Pieterse, M., Bartkova, J., van der, G. H., Hiddingh, S., Thanasoula, M., Kulkarni, A., Yang, Q., Haffty, B. G., Tommiska, J., Blomqvist, C., Drapkin, R., Adams, D. J., Nevanlinna, H., Bartek, J., Tarsounas, M., Ganesan, S., and Jonkers, J. (2010) 53BP1 loss rescues BRCA1 deficiency and is associated with triple-negative and BRCA-mutated breast cancers. *Nat. Struct. Mol. Biol.* 17, 688-695.
53. Bunting, S. F., Callen, E., Wong, N., Chen, H. T., Polato, F., Gunn, A., Bothmer, A., Feldhahn, N., Fernandez-Capetillo, O., Cao, L., Xu, X., Deng, C. X., Finkel, T., Nussenzweig, M., Stark, J. M., and Nussenzweig, A. (2010) 53BP1 inhibits homologous recombination in Brca1-deficient cells by blocking resection of DNA breaks. *Cell* 141, 243-254.
54. Cianfrocca, M., Goldstein, L. J. (2004) Prognostic and predictive factors in early-stage breast cancer. *Oncologist* 9, 606-616.
55. Sun, B., Zhang, S., Zhang, D., Li, Y., Zhao, X., Luo, Y., and Guo, Y. (2008) Identification of metastasis-related proteins and their clinical relevance to triple-negative human breast cancer. *Clin. Cancer Res.* 14, 7050-7059.
56. Cha, S., Imielinski, M. B., Rejtar, T., Richardson, E. A., Thakur, D., Sgroi, D. C., and Karger, B. L. (2010) In situ proteomic analysis of human breast cancer epithelial cells using laser capture microdissection: annotation by protein set enrichment analysis and gene ontology. *Mol. Cell Proteomics* 9, 2529-2544.
57. Mbeunkui, F., Metge, B. J., Shevde, L. A., and Pannell, L. K. (2007) Identification of differentially secreted biomarkers using LC-MS/MS in isogenic cell lines representing a progression of breast cancer. *J. Proteome Res.* 6, 2993-3002.
58. Ou, K., Yu, K., Kesuma, D., Hooi, M., Huang, N., Chen, W., Lee, S. Y., Goh, X. P., Tan, L. K., Liu, J., Soon, S. Y., Bin Abdul, R. S., Putti, T. C., Jikuya, H., Ichikawa, T., Nishimura, O., Salto-Tellez, M., and Tan, P. (2008) Novel breast cancer biomarkers identified by integrative proteomic and gene expression mapping. *J. Proteome Res.* 7, 1518-1528.
59. Hochgrafe, F., Zhang, L., O'Toole, S. A., Browne, B. C., Pinese, M., Porta, C. A., Lehrbach, G. M., Croucher, D. R., Rickwood, D., Boulghourjian, A., Shearer, R., Nair, R., Swarbrick, A., Faratian, D., Mullen, P., Harrison, D. J., Biankin, A. V., Sutherland, R. L., Raftery, M. J., and Daly, R. J. (2010) Tyrosine phosphorylation profiling reveals the signaling network characteristics of Basal breast cancer cells. *Cancer Res.* 70, 9391-9401.
60. Konstantinopoulos, P. A., Spentzos, D., Karlan, B. Y., Taniguchi, T., Fountzilas, E., Francoeur, N., Levine, D. A., and Cannistra, S. A. (2010) Gene expression profile of BRCAness that correlates with responsiveness to chemotherapy and with outcome in patients with epithelial ovarian cancer. *J. Clin. Oncol.* 28, 3555-3561.

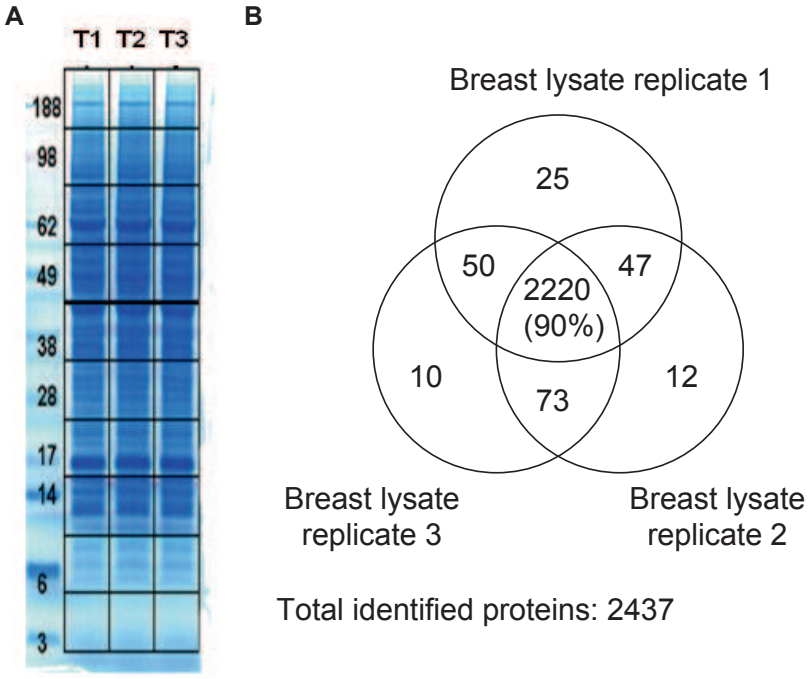
Supplementary Tables

The Supplementary Tables S1-3 can be requested via marcwarmoes@hotmail.com.

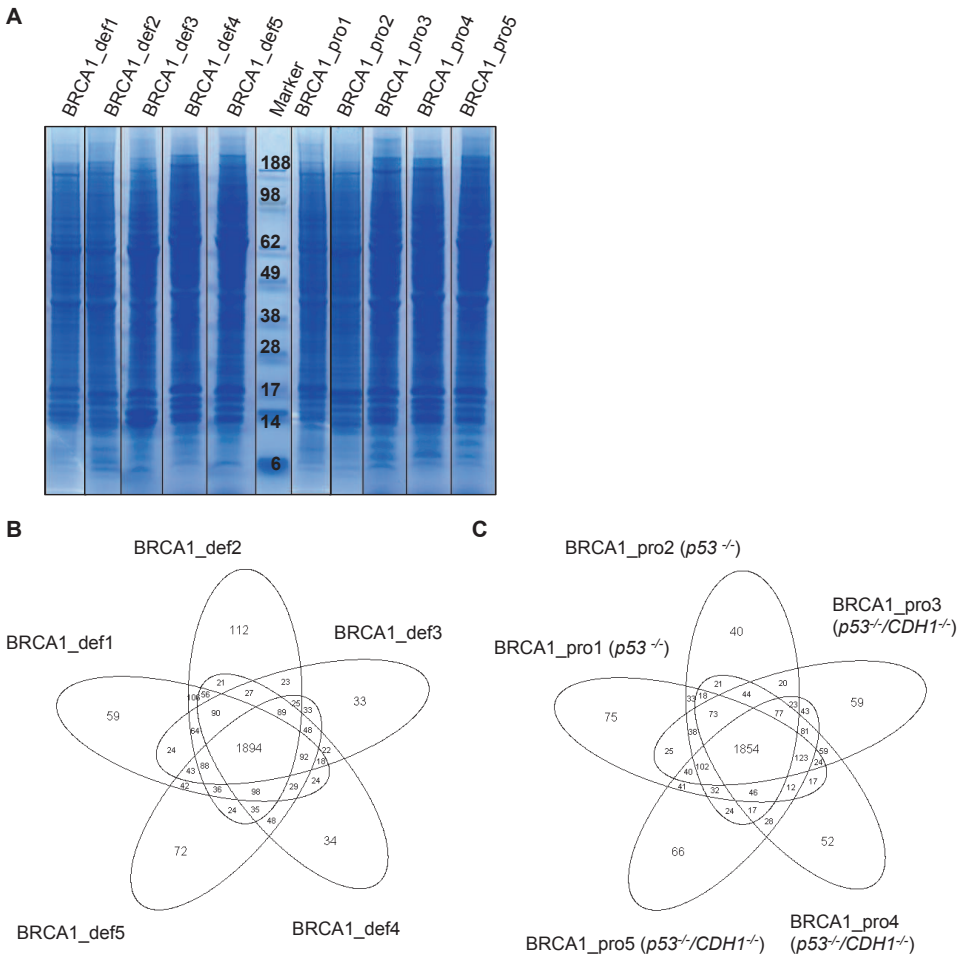
Supplementary Table S1. List of proteins detected in BRCA1-deficient and -proficient mouse tumor tissue lysates and associated spectral count quantification data

Supplementary Table S2. List of significantly differential proteins between BRCA1-deficient and -proficient mouse tumor tissue lysates and associated spectral count quantification data

Supplementary Table S3. List of significantly enriched protein complexes in significantly regulated proteins from BRCA1 deficient mouse tumor tissue lysates

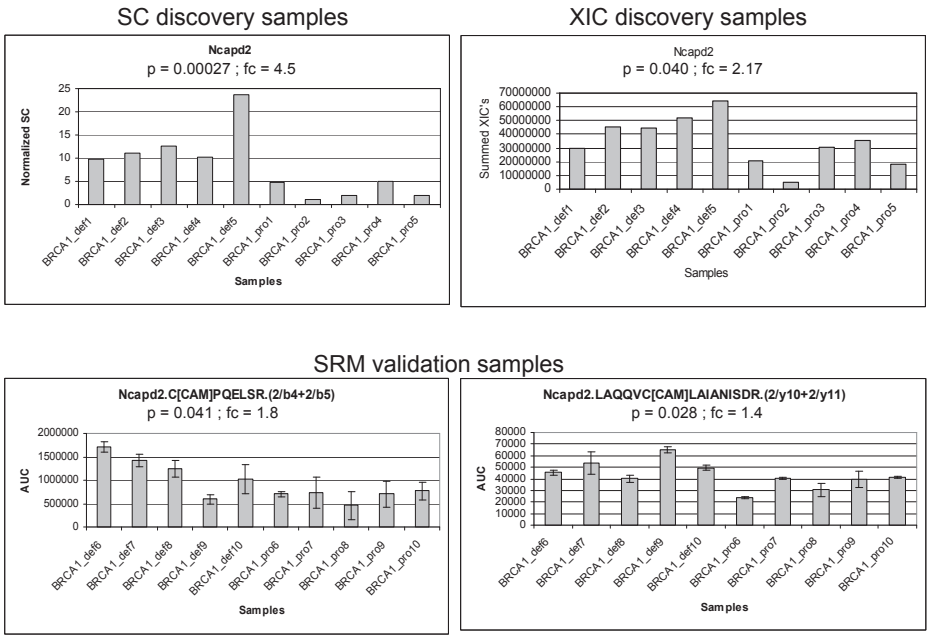


Supplementary Figure S1. Replicate analysis of pooled mammary tumor lysates. A, Coomassie stained gel displaying protein fractionation of three samples from a pooled mammary tumor tissue lysate. **B,** Summary of protein identification by nano-LC-MS/MS. The average CV of the normalized spectral counts is 24 % for the 2220 proteins present in all replicates (90 % of all proteins are overlapping). The total data set contained 2437 proteins.



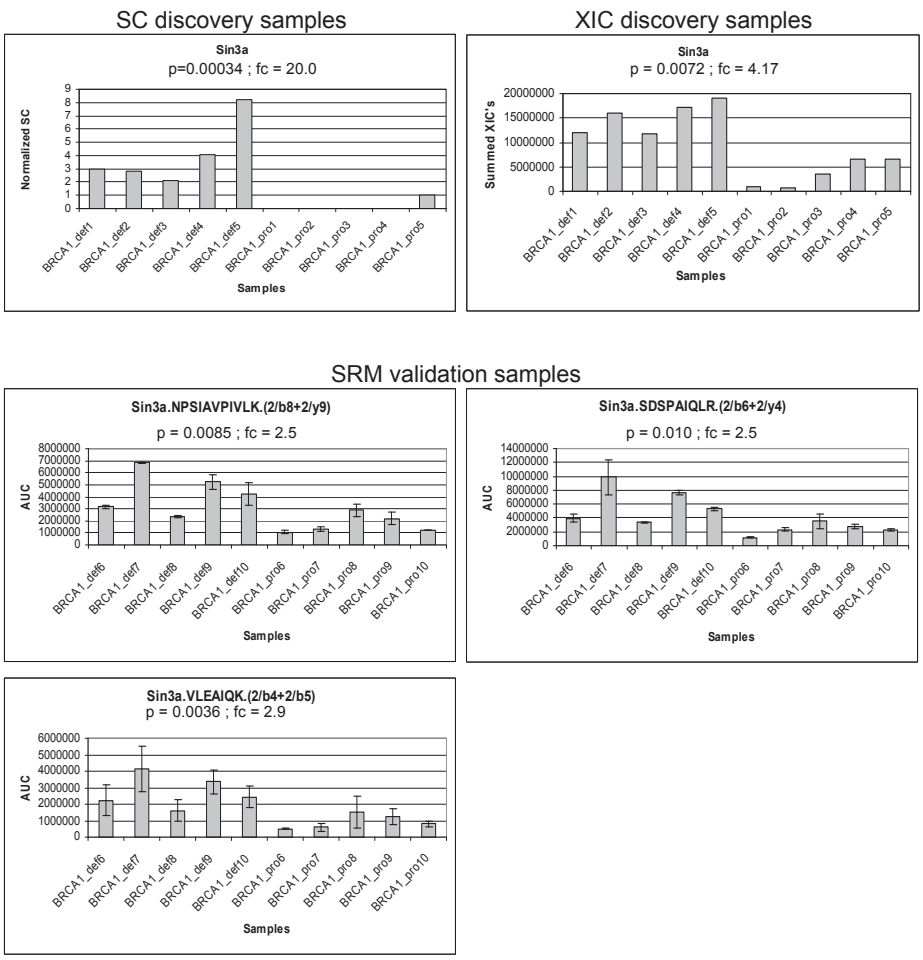
Supplementary Figure S2. Analysis of biological reproducibility. A, Protein fractionation of 10 samples from BRCA1-proficient and -deficient tumor tissue lysates. **B,** Five-way Venn diagram showing the distribution of the protein identifications within the five BRCA1-deficient samples. **C,** Five-way Venn diagram showing the distribution of all 3270 protein identifications within the five BRCA1-proficient samples. 1856 (56 %) proteins were present in all five samples and had an average CV of 36 %.

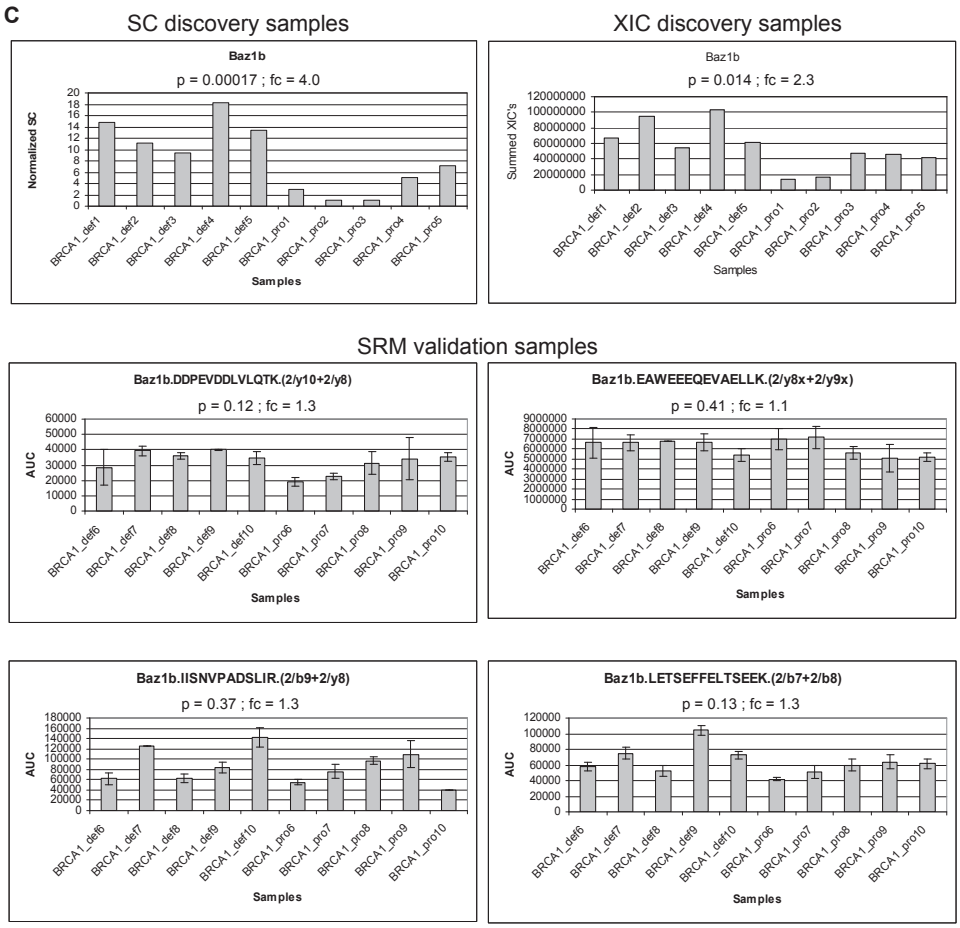
A


SC discovery samples
Ncapd2
p = 0.00027 ; fc = 4.5
Normalized SC
Samples
BRCA1_def1
BRCA1_def2
BRCA1_def3
BRCA1_def4
BRCA1_def5
BRCA1_pro1
BRCA1_pro2
BRCA1_pro3
BRCA1_pro4
BRCA1_pro5
XIC discovery samples
Ncapd2
p = 0.040 ; fc = 2.17
Summed XIC's
Samples
SRM validation samples
Ncapd2.C[CAM]PQELSR.(2/b4+2/b5)
p = 0.041 ; fc = 1.8
AUC
Samples
BRCA1_def6
BRCA1_def7
BRCA1_def8
BRCA1_def9
BRCA1_def10
BRCA1_pro6
BRCA1_pro7
BRCA1_pro8
BRCA1_pro9
BRCA1_pro10
Ncapd2.LAQQVC[CAM]LAIANISDR.(2/y10+2/y11)
p = 0.028 ; fc = 1.4
AUC
Samples

R1
R2
R3
R4
R5
R6
R7
R8
R9
R10
R11
R12
R13
R14
R15
R16
R17
R18
R19
R20
R21
R22
R23
R24
R25
R26
R27
R28
R29
R30
R31
R32
R33
R34
R35
R36
R37
R38
R39

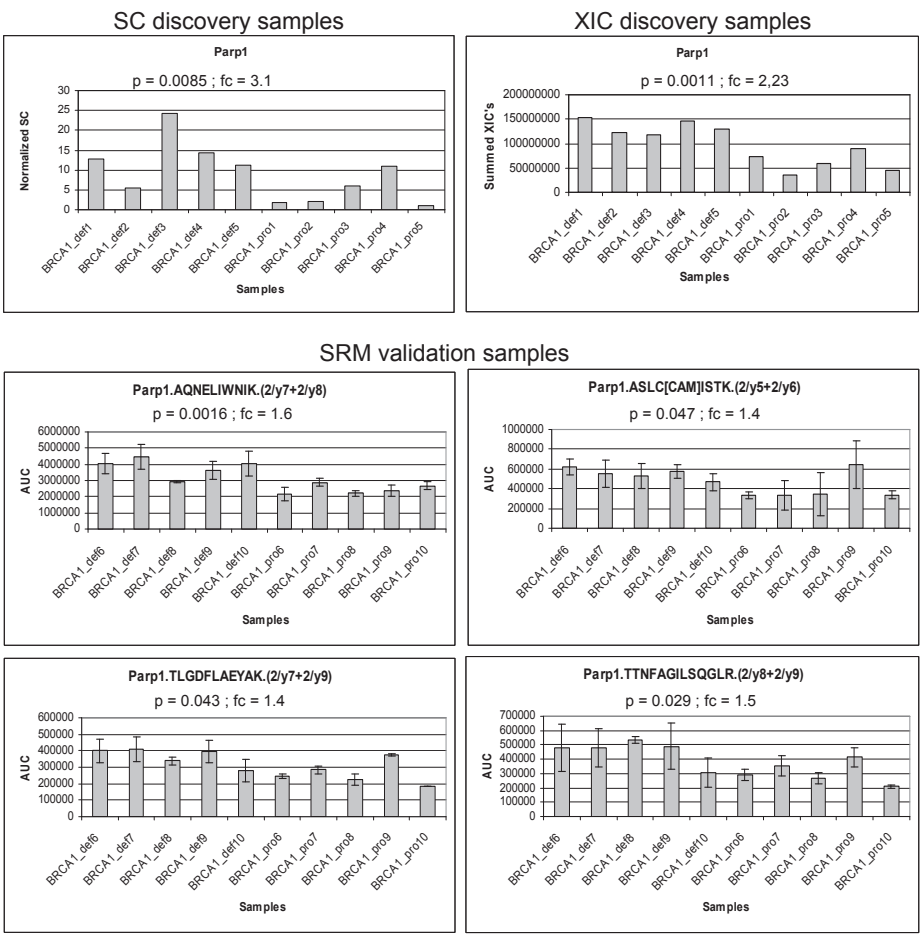
B



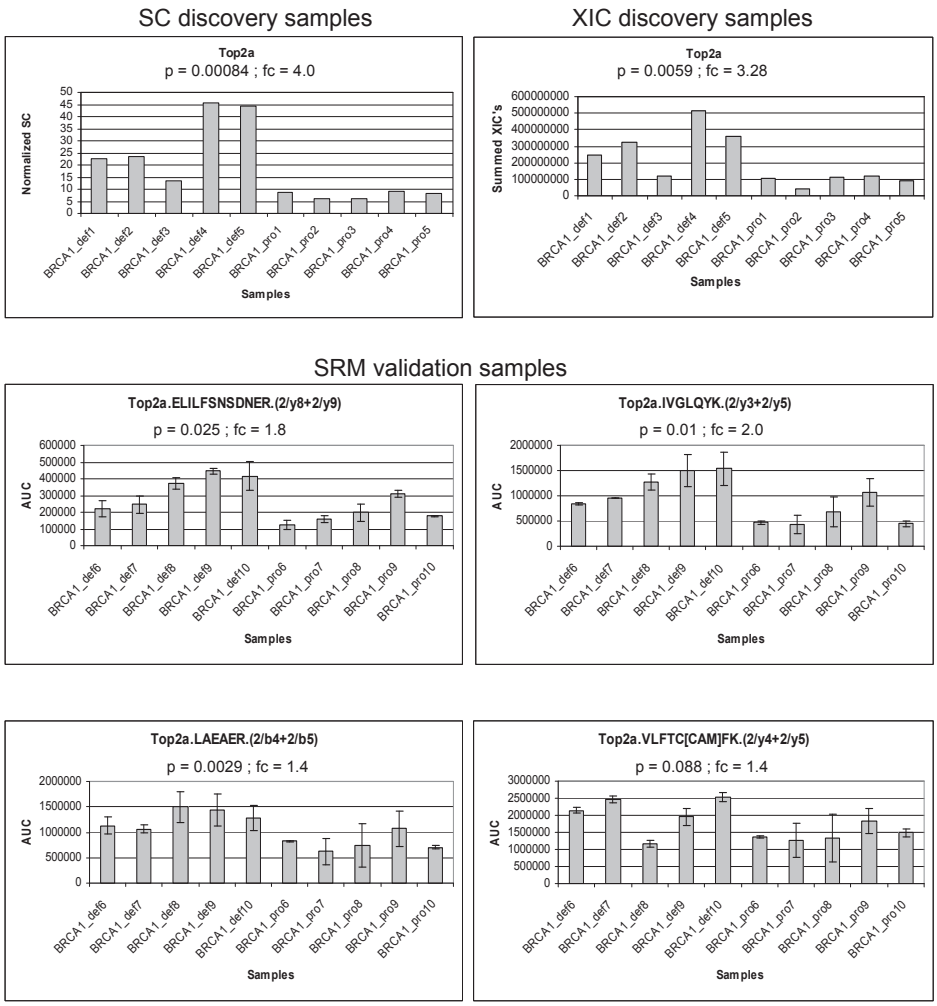

C
SC discovery samples
XIC discovery samples
Baz1b
p = 0.00017 ; fc = 4.0
Normalized SC
Baz1b
p = 0.014 ; fc = 2.3
Summed XIC's
BRCA1_def1
BRCA1_def2
BRCA1_def3
BRCA1_def4
BRCA1_def5
BRCA1_pro1
BRCA1_pro2
BRCA1_pro3
BRCA1_pro4
BRCA1_pro5
Samples
SRM validation samples
Baz1b.DDPEVDDLVLQTK.(2/y10+2/y8)
p = 0.12 ; fc = 1.3
Baz1b.EAWEEEQEVAELLK.(2/y8x+2/y9x)
p = 0.41 ; fc = 1.1
AUC
BRCA1_def6
BRCA1_def7
BRCA1_def8
BRCA1_def9
BRCA1_def10
BRCA1_pro6
BRCA1_pro7
BRCA1_pro8
BRCA1_pro9
BRCA1_pro10
Samples
Baz1b.IISNVPADSLIR.(2/b9+2/y8)
p = 0.37 ; fc = 1.3
Baz1b.LETSEFFELTSEEK.(2/b7+2/b8)
p = 0.13 ; fc = 1.3

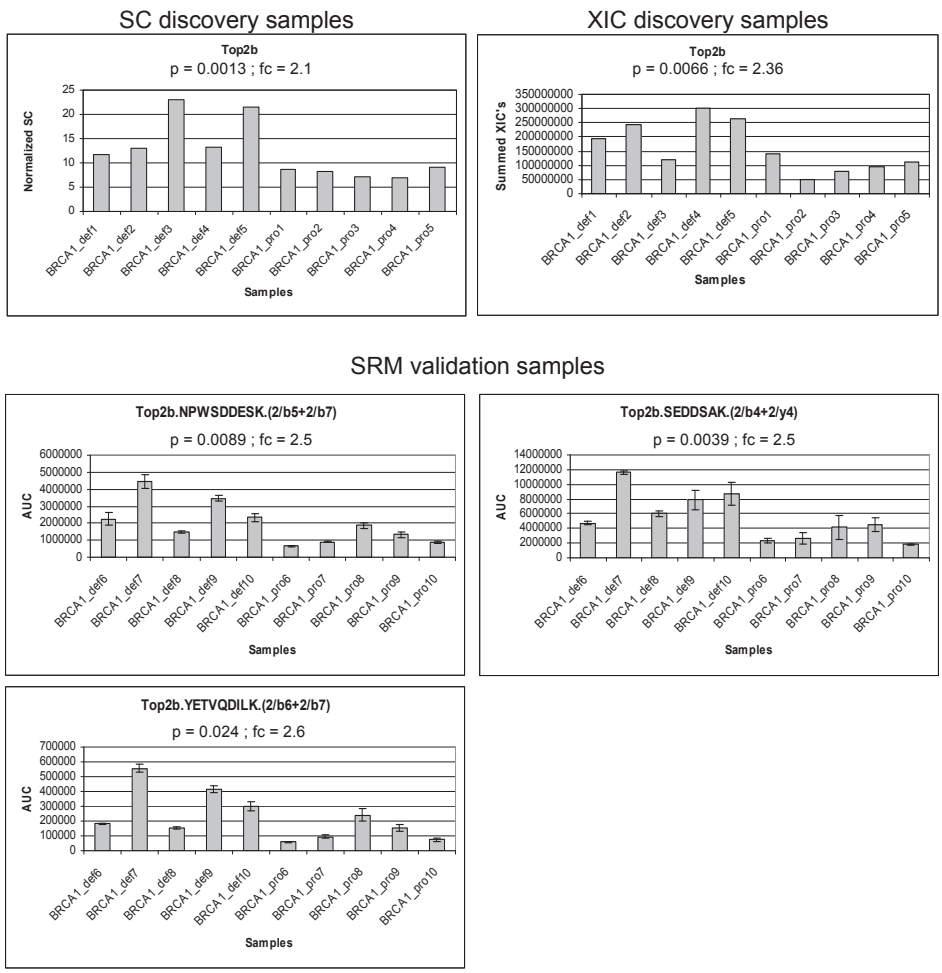
R1
R2
R3
R4
R5
R6
R7
R8
R9
R10
R11
R12
R13
R14
R15
R16
R17
R18
R19
R20
R21
R22
R23
R24
R25
R26
R27
R28
R29
R30
R31
R32
R33
R34
R35
R36
R37
R38
R39

D

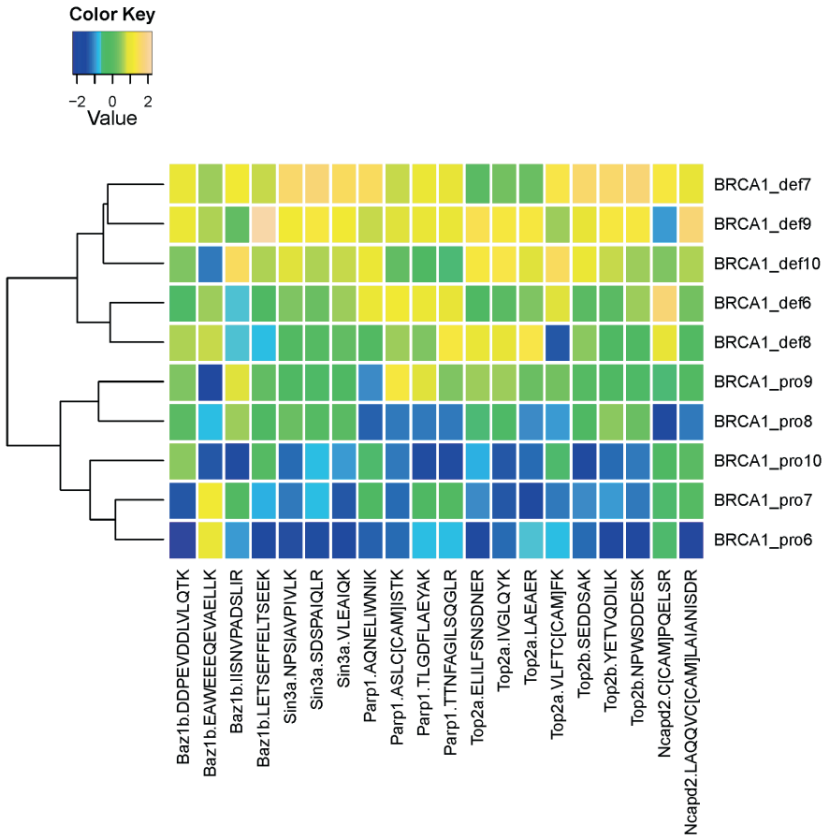


E


SC discovery samples
XIC discovery samples
Top2a
p = 0.00084 ; fc = 4.0
Normalized SC
Top2a
p = 0.0059 ; fc = 3.28
Summed XIC's
BRCA1_def1
BRCA1_def2
BRCA1_def3
BRCA1_def4
BRCA1_def5
BRCA1_pro1
BRCA1_pro2
BRCA1_pro3
BRCA1_pro4
BRCA1_pro5
Samples
SRM validation samples
Top2a.ELILFSNSDNER.(2/y8+2/y9)
p = 0.025 ; fc = 1.8
Top2a.IVGLQYK.(2/y3+2/y5)
p = 0.01 ; fc = 2.0
AUC
BRCA1_def6
BRCA1_def7
BRCA1_def8
BRCA1_def9
BRCA1_def10
BRCA1_pro6
BRCA1_pro7
BRCA1_pro8
BRCA1_pro9
BRCA1_pro10
Samples
Top2a.LAEAER.(2/b4+2/b5)
p = 0.0029 ; fc = 1.4
Top2a.VLFTC[CAM]FK.(2/y4+2/y5)
p = 0.088 ; fc = 1.4



Supplementary Figure S3. Bar graphs representing the normalized area under the curve (AUC) from SRM analysis on four proteins that showed discordant behaviour between protein expression and mRNA expression (NCAPD2, SIN3A, BAZ1B and TOP2B), one protein of which the mRNA level could not be measured because there was no probe on the array (TOP2A), and one positive control that was up-regulated in both the proteomics and transcriptomics measurements (PARP1). Error bars represent the standard deviation of triplicate analysis. **A,** Measurements for NCAPD2; **B,** for SIN3A; **C,** for BAZ1B; **D,** for PARP1; **E,** for TOP2A; **F,** and for TOP2B.



Supplementary Figure S4. Heap map and cluster analysis using peptide intensities in an independent validation set of ten samples (five BRCA1-deficient and five BRCA1-proficient carcinomas). Peptides were derived from proteins that showed discordant behaviour between the protein and mRNA expression. This multiplexed analysis and visualization clearly delineates two groups based on BRCA1 status.